

Measuring Journey Time Reliability in London Using Automated Data Collection Systems

by

Mickaël Schil

Ingénieur Diplômé de l'Ecole Spéciale des Travaux Publics (ESTP)
Paris, France (2011)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

© Massachusetts Institute of Technology 2012. All rights reserved.

Author
Department of Civil and Environmental Engineering
May 21, 2012

Certified by
Nigel H.M. Wilson
Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by
Harilaos N. Koutsopoulos
Visiting Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Measuring Journey Time Reliability in London Using Automated Data Collection Systems

by

Mickaël Schil

Submitted to the Department of Civil and Environmental Engineering
on May 21, 2012, in partial fulfillment of the requirements for the degree of
Master of Science in Transportation

Abstract

Service reliability is critical for both users and operators of transit systems. The rapid spread of Automated Data Collection Systems, such as Automated Fare Collection (AFC) and Automatic Vehicle Location (AVL), provides new sources of information that can be used to measure and assess service reliability.

The main objective of this thesis is to develop a set of simple, customer-driven metrics of journey time reliability, that could be useful and meaningful for both customers and operators. The set of metrics are consistent across transit modes (rail and bus networks).

The proposed methodology, common to rail and bus systems, consists of (1) an analysis of the journey time distributions at the finest spatial and temporal resolution, the origin-destination pair (O-D) and time period level (customer perspective), (2) the aggregation of the reliability metrics at the line (route) level (operator perspective), and (3) the definition of journey time reliability standards at the O-D pair and time period level, by the identification of a representative “good” journey time distribution (both customer and operator perspective).

For fully gated transit systems, like the London Underground, AFC data provides direct travel time measures for every journey from the fare gate at the entry station to the fare gate at the exit station. For non-gated systems, such as many bus networks, no information is available on passengers’ arrival times at the origin bus stop. A method that combines AVL and AFC data is proposed to estimate waiting times at stops so that they can be included in the journey time reliability calculation. Furthermore, the method accounts for the multiple overlapping routes that sometimes serve the same O-D pairs.

The proposed methodology is tested using the London public transport system as an illustration. The use of the reliability metrics for operators and customers is also discussed, with proposed modifications of the information provided by journey planners.

Thesis Supervisor: Nigel H.M. Wilson

Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: Harilaos N. Koutsopoulos

Title: Visiting Professor of Civil and Environmental Engineering

Acknowledgments

I would like to express my deepest gratitude to my advisors, Nigel Wilson, Haris Koutsopoulos, and John Attanucci for their guidance, support and encouragement. Thank you for challenging me and believing in me.

Thank you to Fred Salvucci, Mikel Murga and Juan Carlos Muñoz. Your comments and advice have always been helpful.

Thank you to Ginny and all the CEE staff for your dedication to the students, and for keeping things running in the department and the MST program.

I would like to thank the people at Transport for London for making this research possible. I want to especially thank Duncan Horne, who provided valuable help, suggestions and feedback during and after my internship. Thank you as well to David Glazebrook, Ken Farnhill and Andrew Gaitskell for your help and support over the summer.

Thank you to Gabriel, for your help, advice, support, and for always taking the time to answer my questions and give suggestions. Thanks also for bearing with me as a roommate for this past year. Jay, thank you for your program and the help you provided for using it. Thank you to Alex and Kevin for your help, advice, and above all for these endless conversations about transportation, politics, summer trip plans and other random stuff, that have provided a welcome distraction from work. Naomi, thank you for everything.

Thank you to all the folks in the MST program, especially my officemates in the transit lab. These two years were a lot of fun.

Merci à tous mes amis Parisiens qui, depuis Paris, Nottingham ou Sydney, m'ont soutenu, motivé et remonté le moral. Merci à la French team du MIT pour m'avoir sorti de mon bureau au moins une fois par semaine.

To Ann, Henry, Shaina and Willa, thank you for taking care of me and entertaining me during these two years at MIT.

Finally, I would like to thank my parents who have always pushed me to give the best of myself, and who gave me this opportunity to come to MIT, for two of the hardest, but also two of the best years of my life.

Contents

1	Introduction	15
1.1	Motivation	16
1.1.1	Existing Reliability Metrics	16
1.1.2	Passenger and Operator’s Perspectives	17
1.1.3	Automated Data Collection Systems	18
1.2	Introduction to Transport for London	20
1.3	Research Objectives	21
1.4	Research Approach	22
1.5	Thesis Structure	23
2	Literature Review	25
2.1	Operator’s View of Reliability	27
2.2	Passenger’s Definition of Reliability	28
2.3	Previous Work on Journey Time Reliability Using Automated Data Collec- tion Systems	31
2.3.1	Journey Time Reliability in the London Underground	31
2.3.2	Reliability in the Bus Network	35
2.4	Origin-Destination Inference and Trip Linking	36
2.5	Conclusion	37
3	Current Reliability Metrics used by Transport for London	39
3.1	London Underground	39
3.1.1	Journey Time Metric	42
3.1.2	Lost Customer Hours	48
3.2	London Buses	49

3.2.1	Excess Wait Time	51
3.2.2	Percentage Chance of Waiting More Than a Given Threshold	53
3.2.3	Publication of the Metrics	53
3.3	London Streets	54
3.4	Conclusions on the Current Reliability Metrics	56
4	Measuring Journey Time Reliability in the London Underground	57
4.1	Requirements for the Journey Time Reliability Metrics	62
4.2	Reliability Measurement Framework for a Single Underground Line	63
4.2.1	Data Available	63
4.2.2	Origin-Destination Pair and Timeband Analysis	64
4.2.3	Aggregation at the Line Level	67
4.2.4	Definition of the Journey Time Reliability Standards	72
4.3	Journey Time Reliability Measurement for Comparison Across Lines	77
4.4	Applications	79
4.4.1	Data Processing	79
4.4.2	Journey Time Reliability Metrics at the Origin-Destination Pair Level	80
4.4.3	Journey Time Reliability Metrics at the Line Level	85
4.4.4	Journey Time Reliability Standards	87
4.4.5	Comparison of the Daily Performance with a Rolling Average Median Journey Time	94
4.4.6	Metric for Comparison of Underground Lines	96
4.5	Use of the Reliability Metrics	98
4.5.1	Presentation for Internal Use at Transport for London	98
4.5.2	Presentation to the Customers: Addition to the Journey Planner . .	103
4.6	Conclusion - Assessment of the Proposed Metrics	106
5	Measuring Journey Time Reliability for High Frequency Bus Services	109
5.1	Assumptions	110
5.2	Methodology at the Origin-Destination Pair Level	113
5.2.1	Data	113
5.2.2	Waiting Time Estimation	114
5.2.3	In-Vehicle Travel Time	118

5.2.4	Total Journey Time	118
5.2.5	Waiting Time and Journey Time Reliability Standards	121
5.3	Aggregation at the Route Level	122
5.4	Applications	124
5.4.1	Data Processing	125
5.4.2	Results at the Origin-Destination Pair Level	126
5.4.3	Results at the Route Level	137
5.5	Use of the Reliability Metric	138
5.5.1	Internal Use at London Buses	138
5.5.2	Journey Planner Modification	138
5.6	Conclusion - Assessment of the Proposed Metrics	141
6	Conclusion	145
6.1	Summary	145
6.2	Findings and Recommendations	146
6.3	Future Research	148
6.3.1	Fully Gated Systems	148
6.3.2	Bus Networks	149
6.3.3	Extension of the Metrics to Multimodal Journeys	150
A	Tube Map	153
B	London Underground Reporting Periods	155
C	London Underground Daily Dashboard	157
D	Route 141 Northbound	159

List of Figures

2-1	The likelihood of delays as a function of the RBT upper percentile value . .	32
2-2	Classification into recurrent and incident-related performance	33
2-3	Definition of recurrent performance	34
3-1	Contribution of each journey component to Total Journey Time	45
3-2	Weighed Excess Journey Time Results - Period 9, 2011/12	46
3-3	Lost Customer Hours - Period 9, 2011/12	49
3-4	Route 141 - Map and QSI points	50
3-5	Excess Wait Time for high frequency bus services - Long term trend	54
3-6	Example of determining which journey times are acceptable / unacceptable	55
3-7	London Streets Journey Time Reliability - Quarter 3, 2011/12	56
4-1	Map of the 3 origin-destination pairs	58
4-2	Journey planner - Canary Wharf to Bond Street	59
4-3	Journey time distributions for 3 O-D pairs - PM peak	60
4-4	Margin of error of the Oyster journey time - Individual journey	64
4-5	Margin of error of the Oyster journey time - Simultaneous trips	65
4-6	Illustrative journey time distributions	67
4-7	Finsbury Park to Green Park	71
4-8	Journey time reliability metrics - Jubilee line	85
4-9	Reliability Buffer Time distributions - Periods 7, 8 and 9 - Weekdays	88
4-10	Comparison of actual and good journey time reliability metrics - Three O-D pairs	91
4-11	Comparison of actual and good journey time reliability metrics - Jubilee Line	93

4-12 Comparison of the daily performance with a rolling average median journey time	95
4-13 Journey time reliability metrics: Jubilee line vs. Metropolitan line	97
4-14 Comparison of the Reliability Buffer Time with the Platform Wait Time Proxy	100
4-15 Comparison of the Percentage of journeys over the good journey time + 5 minutes with the Lost Customer Hours	101
4-16 Correlation between the existing and the proposed metrics	102
4-17 Proposed modification of the journey planner - Direct trip	103
4-18 Proposed modification of the journey planner - Journey with an interchange	105
5-1 Bus routes from Bank Station	111
5-2 Headways and bus arrivals	115
5-3 Passenger arrival probability density function	116
5-4 Passenger journey time probability density function	119
5-5 Route 141 map as of November 2011 with boroughs and rail stations	124
5-6 Time-space diagram of successive bus trips	126
5-7 Output headways and running times	126
5-8 Headway cumulative distribution for destinations from the Bank Station / Princes Street stop	128
5-9 Waiting time cumulative distributions for destinations from the Bank Station / Princes Street stop	129
5-10 Waiting Time, In-vehicle Travel Time, and Journey Time distributions . . .	132
5-11 Journey time distributions for 4 O-D Pairs - PM Peak	133
5-12 Reliability metrics - Bank Station / Princes Street to Wood Green Station - PM peak	135
5-13 Aggregate Reliability Metrics - Route 141 Northbound - PM peak	137
5-14 Journey Planner - Bank Station / Princes Street to Mintern Street	139
5-15 Proposed modification of the Journey Planner for bus trips	140
5-16 Stop areas	142

List of Tables

2-1	Five groups of metrics identified by the customers	29
3-1	Key Performance Indicators Evaluation	41
3-2	VOT weights for JTM calculation	45
3-3	London Buses Network Performance - Third Quarter 2011/12	53
4-1	Example of processed Oyster records	80
4-2	Journey time reliability metrics for 3 O-D pairs	81
4-3	Seasonality effects - Canary Wharf to Bond Street - PM peak	83
4-4	Seasonality effects - Heathrow Terminals 1,2,3 to Hammersmith - PM peak	84
4-5	Seasonality effects - Jubilee line	86
4-6	Correlation coefficients between the exiting and the proposed metrics	99
5-1	Data sources and uses	114
5-2	Waiting time reliability metrics for destinations from Bank Station / Princes Street	131
5-3	Journey time reliability metrics for destinations from Bank Station / Princes Street	136
6-1	Four interchange cases	151
D-1	Route 141 Northbound - List of stops and parallel routes	160

Chapter 1

Introduction

Reliability is defined as the quality of being consistently good in quality or performance, and the ability to be trusted¹. Journey time reliability is critical for people using public transportation systems. In order to plan their trips, passengers need to know how long their journeys should take. If they make the same trip on a regular basis, passengers prefer that their journey takes about the same time every day. It is also important for operators for their service planning, crew and vehicle scheduling, and service contract design.

Vehicle on-time arrival and/or headways are often used by public transport agencies as measures of reliability. These metrics typically compare actual operations with the schedules. Knowledge that 95% of the trains on a specific line arrive on time at the terminal does not necessarily give passengers information about their own journeys defined by specific origin-destination (O-D) pairs. Traditionally, reliability metrics reported by transit agencies are calculated from an operations perspective and do not always relate to the users' own journeys and experiences

The rapid spread of Automated Data Collection Systems (ADCS), including fare collection and vehicle location systems, provides precise and detailed data on running times and passenger travel times. In most case, the data covers almost all vehicles operating and all passenger trips. ADCS also have the potential to provide more customer-oriented information. However, this data are not currently used to its full potential by transit providers. The direct data provided by ADCS is still raw and needs processing.

The purpose of this thesis is to extend prior research by Chan (2007), Uniman (2009)

¹source: Oxford Dictionaries

and Ehrlich (2010) on journey time reliability metrics, in order to design and test reliability measures consistent across the various modes that are operated in parallel in many transit systems. The metrics will be based on automated fare collection and automatic bus vehicle location systems, as both sources provide abundant information for this purpose. The London public transport system is used as an application where the proposed metrics are evaluated. The automated data collected by the agency Transport for London (TfL) provides a great platform for this research.

While this thesis focuses on London, the methodology presented here can be applied to any other transit agency that collects similar data.

1.1 Motivation

1.1.1 Existing Reliability Metrics

Many transit systems include more than one mode (e.g. bus, rail, etc.). In these systems, passengers often use more than one mode in a single journey. One could, for example, use a bus from home to reach the commuter rail station, then take a train to the city center before transferring to the subway to reach the work place. Therefore, passengers are more interested in reliability metrics that cover their entire trip, and in seeing similar metrics for all the modes they use. However, transit agencies often report reliability metrics separately for each mode. For instance, the Paris transportation authority, the “Syndicat des Transports d’Île de France” (STIF), reports and publishes different metrics for each of the modes operated in the Paris region (STIF, 2012).

- For the metro, the average percentage of customers waiting less than 3 minutes in the peaks, 6 minutes off-peak, or 10 minutes in the evening is calculated for each of the subway lines for each quarter.
- For the bus network as a whole and for each light rail line, the reliability metric compares the actual headways with the schedules at major stops on each line.
- For the “Réseau Express Régional” (RER) and suburban rail lines, the scorecard shows the percentage of passengers that arrive at their destination with a delay no greater than 5 minutes.

The Massachusetts Bay Transportation Authority (MBTA), which operates the public transportation system in the greater Boston region, periodically reports on-time performance, but only for the rail network (MBTA, 2012).

- For the subway lines, the performance metric compares the actual frequencies to the scheduled ones.
- For the commuter rail lines, the metric calculates the percentage of trains that arrive with a delay no greater than 4 minutes.
- No travel time reliability measure for journeys on the bus network is reported to the customers.

Customers are generally not provided with a metric that reflects either their entire journey or their own experience. The reliability metric for the Paris metro only takes into account the platform wait time, and does not consider the time spent in the train. One could argue that time spent waiting on the platform is weighted more by the passengers than time spent in the vehicle; however the unreliability of the in-vehicle travel time can be as inconvenient as the waiting time. Moreover, for multimodal journeys that combine more than one mode, there is no single metric that evaluates the performance of the trip as a whole. A good metric for travel time reliability should take into account the entire journey experience, including any interchanges.

1.1.2 Passenger and Operator’s Perspectives

Passengers and operators have similar objectives as they all want travel times in the public transportation system to be reliable. However, their perspectives are different. Passengers have interest in their own journeys, expecting on-time arrival and reliable travel time. Reliability metrics will help them plan their future journeys. The operator on the other hand is more concerned with the performance of each rail line or bus route, or of the system as a whole.

The current reliability metrics give results at the line or network level, and do not provide daily information for specific origin-destination pairs. The published measures typically show average values for a line or route over a period of a month or a quarter. These results are useful for the operators or the transit agencies as they attempt to capture

the performance of the transit system into aggregated values. These indexes can be used for service monitoring and comparison of performance among lines. This is also particularly useful in the case where service is contracted out and the contract includes performance-based penalties or incentives. Given the results, actions are then taken to improve the quality of service if the performance objectives were not reached (compared to target values).

Passengers have less interest in the performance of a subway line or a bus route as a whole as they are more concerned with the journey time reliability of the trip between their origin and their final destination. They do not relate the metrics they are given to their own journeys. Knowing that on line 1 of the Paris metro, 99% of the passengers waited 3 minutes or less on average in the peak periods between January and September 2011 (STIF, 2012), does not give customers information about actual waiting times for their past or future trips. The travel conditions of the remaining 1% is not mentioned either. Additionally, customers who are unfamiliar with the transit network often use online journey planners, that are becoming increasingly available. Users of the system tend to rely on the information given by the journey planner, at least initially. Journey planners do not give any indication of travel time variability, as the journey times published are based on schedules. Only passengers who frequently travel between the same origin and destination have, from experience, an idea of the travel time variability, and plan their trip accordingly.

Despite the passengers' and operators' common objectives, the travel time reliability metrics are too operator-focused, and no measure today translates the reliability performance of the transit lines into numbers useful for the customers' own journeys. A good metric for passengers would provide information on journey time variability at the O-D pair level (to capture the entire journey experience), not just averages at the line (route) level over several weeks. A good metric would also help customers plan their future journeys.

1.1.3 Automated Data Collection Systems

The current service reliability performance indicators are mainly based on manual surveys and models to determine waiting and/or travel times, and only report an average journey experience. Manual data collection is time consuming and expensive, and the results may not be accurate due to sampling, measurement, and other errors.

The increasing implementation of Automated Data Collection Systems in transit networks around the world, including Automatic Vehicle Location (AVL), and Automated Fare

Collection (AFC) systems, facilitates the collection of data for the analysis of the performance of transit systems in cost-efficient and effective ways. These sources of data can be used to develop metrics for travel time reliability with precise information at fine levels of granularity, by origin-destination, and time of day. This level of detail cannot be accomplished by manual methods.

AVL systems track and record bus or train location at all or specific stops along the route. AVL thus provides exact running times between the recording points for all trips. AFC systems record useful information that can be used for performance analysis. In the Paris region, the smart card “Passe Navigo” has to be validated when boarding the bus, entering the metro, transferring from the metro to the RER, and when entering or exiting the RER. In Boston, customers using the “Charlie Card” have to validate their card when entering a station or boarding a bus or light rail train. Regardless of the validation pattern, AFC provides useful data, especially when combined with information from other automated sources such as AVL.

For systems with both entry and exit validation, the AFC data provides the exact travel time between the origin station fare gate and the destination station fare gate. For systems that require validation only when the passenger enters the rail system or boards the bus, the in-vehicle travel time cannot be calculated using only AFC data; AVL information is needed to measure travel times. Furthermore, the destination of the passenger is not known with certainty, and can only be inferred from the following validation of the fare card. Gordon (2012) presents a framework for destination inference for bus journeys, and results of his research will be used in this thesis. Finally, waiting times for bus journeys cannot always be calculated from the automatically collected data. The arrival time of the passenger at the bus stop is not recorded, and therefore the waiting time until the bus arrival is not known. In these cases, estimation of the passenger waiting time at a stop is needed in order to be able to estimate journey times for bus trips.

While the automated data systems do not allow the breaking of journeys into their different parts, they provide a granularity in the data that allows calculation of travel time from every origin to every destination in the system, for any time period of the day. The results can then be aggregated at the appropriate spatial and temporal levels, for instance at the line (route) and day level, to provide metrics that are more relevant to the operator.

1.2 Introduction to Transport for London

The London public transportation system includes a large variety of modes: London Buses with more than 700 routes and a fleet of about 8,500 vehicles, the London Underground with 11 subway lines, the London Overground, a circumferential suburban rail network, the Docklands Light Rail (DLR), a light metro that operates in the east of the Capital, the London Tramlink in Croydon and the London River Services on the Thames (Transport for London, 2010). Since 2000, Transport for London (TfL) has been in charge of managing the transportation system in the Greater London area. The modes are operated independently by different subsidiaries within Transport for London.

Similarly to Paris or Boston, each mode focuses on its own performance and so service reliability is currently reported using different measures across the modes. These metrics even include different journey time components. London Buses reports Excess Wait Time (EWT) for high frequency services, estimating the average waiting time at the stop and comparing it with the average scheduled waiting time. London Underground calculates the Journey Time Metric (JTM) and reports the Excess Journey Time (EJT). These metrics use the journey time from the origin to destination. Each stage of the journey is weighted by the value of time. The actual journey time is also compared with the scheduled time. London Overground reports the Public Performance Metric (PPM), a measure of on-time performance at the terminal, defined as the percentage of trains that arrive no more than 5 minutes after their scheduled arrival time.

Several automated data collection systems have been implemented by Transport for London. Only the Oyster card and the iBus systems are used in this thesis. The Oyster smart card fare collection system was introduced in 2003. It can be used for passes or pay-as-you-go transactions for journeys in the London public transport system, on modes including buses, Underground, Overground, and the National rail in Greater London (Transport for London, 2012d). The system records all the transactions, including (most importantly) the entry (boarding) times, as well as the exit times in the rail network. As of April 2012, more than 80% of all bus and Underground transactions are made using an Oyster Card.

iBus is London's real-time bus location system (Transport for London, 2012a). This AVL system was developed starting in 2003 to replace the old Band III radio system (Ehrlich, 2010), and was progressively installed on the entire network from 2007 to 2009. iBus uses

GPS to track the location of all the vehicles on the network. The system records the arrival and departure of each bus at each stop, and helps the operator decide on the best strategies for service control. iBus data is also used to provide real-time information on bus arrivals displayed on the countdown signals at bus stops, and for on-board next stop announcements.

1.3 Research Objectives

The objective of this thesis is to develop and test a methodology to measure journey time reliability in a transit system where several modes are operated in parallel. The proposed metrics should be consistent across modes, and should include the same components of the journey time. The metrics should reflect the entire passenger's experience from his origin to his destination, and including interchanges between vehicles and modes. Finally, the metrics should be customer-focused, but should also be relevant for the operators. The main objectives of the thesis are to:

1. Define journey time reliability and its meaning for customers and operators.
2. Propose a set of metrics to evaluate journey time reliability in a similar fashion for all transit modes. The metrics should reflect customers' experience by capturing the entire journey time, and considering parallel routes that can serve the same origin-destination pairs. The new measures should also complement existing metrics by offering a finer level of granularity.
3. Define journey time reliability standards based on customers' actual experience and use these standards for evaluation of the daily performance of the system.
4. Test the proposed metrics using AFC and AVL data from the London transit system.
5. Propose presentations of the metrics, understandable to both the customers and the operator, aggregated in time and space in an appropriate fashion.

1.4 Research Approach

The general approach is based on the analysis of travel time distributions, considering origin-destination pairs independently, for each days and each time period within a day. Calculation of line-level metrics is then considered. A methodology to define journey time reliability standards at the origin-destination pair and timeband levels is also proposed, in order to define a basis for comparison of daily performance that is not based on schedules.

Travel time reliability framework for fully gated systems. The definition and test of reliability measures for fully gated systems, such as the London Underground, are the first steps of this thesis. AFC data for journeys made in a fully gated system, with entry and exit validations, provides the exact journey time for each trip made in the system. This disaggregate data supports an analysis at the origin-destination pair level, for every day and every timeband. Metrics will be proposed and discussed, and reliability standards established.

Extension of the metrics to open-gated systems. The thesis will then consider the extension to open-gated systems like bus networks, using AFC and AVL data. In the case of open-gated system, AFC data alone does not provide journey time information since cards are only validated upon entrance in the system but not when leaving ths system. In order to define metrics comparable to the ones defined for fully gated systems, a method for estimation of the waiting times at each bus stop, according to the passenger's destination is required. The methodology should also take into account that in the case of multiple routes serving the same O-D pair, passengers waiting time is based on the combined headways (among all relevant routes). The total journey time is estimated as the convolution of the estimated waiting time and the known in-vehicle travel time for individual trips. Aggregation and reliability standards are also discussed.

1.5 Thesis Structure

Chapter 2 presents a reviews of the literature related to service reliability and the way customers perceive it. Previous theses discussing reliability for rail and surface transport and introducing the concept of reliability buffer time are also reviewed. The work by Gordon (2012) on origin-destination inference and trip chaining is also introduced. Chapter 3 presents and analyzes the different service reliability metrics currently used by the different businesses at Transport for London.

Chapter 4 proposes a framework for alternative reliability metrics based on travel time distributions for subway trips, using London Underground Oyster records as a data source for this analysis. Chapter 5 extends the reliability metrics to the bus network, by proposing a method for estimating the waiting time distribution at the origin destination pair level for each timeband. Chapter 6 summarizes the research, concludes with a discussion of outcomes, strengths and weaknesses of the metrics, and provides suggestions for future work.

Chapter 2

Literature Review

The notion of *service reliability* is at the core of this thesis. Reliability encompasses different aspects, and has important implication for transit operators and transit users.

Abkowitz et al. (1978) define reliability in transportation as “*the invariability of service attributes which influence the decision of travelers and transportation providers*”. Lomax et al. (2003) distinguish the concept of *reliability* and *variability* for their difference in focus and measurement. *Reliability* refers to the “*level of consistency in transportation service for a mode, trip, route or corridor for a time period*”, and relates to customers’ experience and perception of their trip. *Variability* is operator-focused, and is defined as “*the amount of inconsistency in operating conditions*”. Carrion and Levinson (2012) state that travel time variability is caused in general by *predicable* (e.g. peak hour congestion) and *unpredictable* (e.g. incidents) variations. Travel time reliability more specifically is linked to the unpredictable variations. From these three considerations two key ideas emerge:

1. Operators and customers perceive reliability differently, and an ideal measure of reliability would encompass the two perspectives;
2. An appropriate measure of travel time reliability should capture the impact of unpredictable events on customers and operations.

As highlighted by Furth and Muller (2006), traditional measures of reliability only provide mean values while passengers usually remember extreme conditions. Mean values translate into operational performance rather than reliability’s impact on customers. There has been a large amount of research conducted on travel time reliability. Over the past

twenty years, with the emergence of Automated Data Collection Systems (ADCS), the perspective of reliability measurement has shifted from operator-focused to customer-focused. Furth et al. (2006) explain how the process of implementing full scale ADCS, like Automatic Vehicle Location (AVL) or Automatic Passenger Count (APC) has allowed transit agencies to shift from being “data poor to data rich”. The availability of data from ADCS allows transit agencies to focus on (Furth and Muller, 2006):

- *Extreme values.* Manual surveys used to support the traditional reliability metrics only provide small samples and average values. The large samples provided by automatically collected data allow the calculation of distributions of quantities of interest (e.g. headways, journey times) and the estimation of percentiles that capture the worst performances (e.g. 95th percentile).
- *Customer-oriented service standards and scheduling.* AVL and APC data allow the definition of metrics that capture customers’ experiences when they travel. Furth et al. (2006) suggest for instance the use of the percentage of passengers waiting longer than x minutes, where x is “a threshold for unacceptability”, as a reliability standard.
- *Planning for operational control.* AVL data allows the measurement of the effects of real-time control decisions, such as curtailments on a bus route.
- *Solutions to roadway congestion.* AVL data can be used to quantify the effect of traffic management schemes and signal priority.
- *Discovery of hidden trends.* New trends previously considered as random variations can be identified with the large samples provided by AVL and APC data.

This thesis aims at two of these areas: the opportunity to understand better the impact of extreme values and their use in developing reliability metrics, and the new customer-oriented service standards.

Lomax et al. (2003) identify three types of potential metrics that focus on the passenger’s perspective on reliability:

- *Statistical ranges.* These measures evaluate the dispersion of the travel conditions experienced by passengers around a central value like the mean or the median.

- *Buffer time measures.* Buffer times refer to the additional time that a customer must budget in order to arrive on time at the destination. Buffer times can be measured as a percentage of the average trip time or as an absolute value.
- *Tardy trip indicators.* These measures evaluate the likelihood of extremely long, or “unacceptable” delays. A threshold between acceptable and unacceptable lateness is defined, either as a percentage of the “normal” journey time or as an absolute value.

These three metrics are all related to the two opportunities mentioned above, the focus on extreme values, and the customer-oriented metrics.

This chapter summarizes the operator’s and passenger’s perspectives on reliability and describes previous works on journey time reliability in London by Chan (2007), Uniman (2009), and Ehrlich (2010). Section 2.1 briefly reviews traditional operation-focused reliability metrics. Section 2.2 proposes some definitions of reliability from the customer’s point of view. Section 2.3 summarizes previous work on the use of AVL and AFC data to measure journey time reliability. Finally, section 2.4 describes the work by Gordon (2012) that, although not directly related to reliability, is a critical input for this thesis.

2.1 Operator’s View of Reliability

Traditional operation-based reliability measures, summarized in Uniman (2009) and Trompet et al. (2011), include:

- *Headway regularity* is measured for high frequency services (usually defined as headways of twelve minutes or less), for which passengers are assumed to arrive randomly. Different metrics can be used to evaluate headway regularity, for instance the standard deviation of the difference between the actual and scheduled headways, or the percentage of headways that deviate no more than a specific amount from the scheduled headway (Trompet et al., 2011).
- *Schedule adherence* is used to measure reliability performance for low frequency service. It is measured as the percentage of vehicles that depart or arrive within an acceptable time interval around the scheduled departure or arrival time.

- *Average passenger waiting time* is the most commonly used metric for service reliability, calculated as follows:

$$E(w) = \frac{E(h)}{2} [1 + cov^2(h)] \quad (2.1)$$

where $E(h)$ is the expected headway and $cov(h)$ is the coefficient of variation of headways. The average waiting time is often compared to the average scheduled waiting time to calculate the Excess Wait Time. These waiting time metrics are based on operational data; however, they are considered as the most representative of passenger experience.

These three traditional metrics measure average service reliability with a focus on operational performance rather than passengers' experience. Before reviewing previous work on passenger-focused reliability metrics, the following section defines passenger's perception of reliability, and what they expect from reliability metrics.

2.2 Passenger's Definition of Reliability

The definition of reliability from the customer point of view is the starting point of this thesis. London Underground has conducted an extensive study (Transport for London, 2011a) with the following objectives:

- understand how people experience and perceive reliability in the Underground;
- identify reliability metrics that would be meaningful and useful for the customers;
- test customer reactions to several existing and potential metrics.

Customers were asked about reliability in general, and not specifically about travel time reliability. Two definitions given by London Underground customers are proposed below:

"A reliable someone/something is dependable and trustworthy that leads to a pleasant experience"

"Reliable is... something that's dependable, trustworthy, honest, reassuring, and something you can count on."

Predictability and *dependability* are the two key ideas that emerge from these definitions. The “*pleasantness*” is also considered as an important part of reliability in the Underground.

An important finding of this study is that customers value personal, “me-centric” information and metrics. Line and system-wide averages, as currently provided by Transport for London, are generally not good enough for customers. Average metrics do not necessarily relate to their own experiences. Reliability must also be considered at the whole journey level, and all the attributes (e.g. PWT, OTT, passenger information, safety) of the journey must be taken into account.

Two sides of reliability are highlighted by customers: the operational reliability, and the customer care reliability. The two aspects cannot be considered independently. Operational reliability focuses on the service provided by TfL, in terms of service frequency, wait time, comfort, predictability; customer care reliability relates to communication and help to passengers, especially in the event of disruptions and other abnormal situations. The help provided for journey planning is part of the customer care aspect.

Customers were introduced a range of existing and potential metrics, and asked to evaluate them in terms of relevance, credibility, and interest. They could also suggest new metrics. Customers classified the metrics in five different groups, as shown in Table 2-1.

Operational	Negative	Retrospective	Big Facts	Future Focused
- 97% of lifts in service - 96% of escalators in service - 99% of trains in service - 98% km operated	- Number of delays on train that are more than 5/10/15/30 minutes - Number of detrainments - Number of trains stuck between stations for more than 30 minutes - 48 out of 260 stations for more than 15 minutes	- 6 minutes excess journey time on average - 44 minutes total journey time on average - 79 out 100 overall score on customer satisfaction survey - % of journeys taking 5/10/15/30 minutes more than they should - % of journeys where passengers wait on a platform for 5 minutes or more than they should	- Top causes of delays - 1,065 million total number of journeys carried each day	- Train frequency - Platform wait time - averages across the network, line, station, at peak and off peak - # of signal failures

Table 2-1: Five groups of metrics identified by the customers

Customers reacted differently to each group of metrics:

- *Operational.* Metrics such as the percentages of lifts or trains in service are not well interpreted by the customers. They give system-wide information that does not necessarily correlate with customers' personal experience.
- *Negative.* Negative metrics such as the number of detrainments or the number of delays give a bad reputation to TfL and the Underground. These metrics should not be publicized, especially given that media tend to highlight negative performance and failures.
- *Retrospective.* Retrospective average statistics are not relevant to customers, as customers automatically compare the retrospective metrics with their personal experiences. They tend to invalidate the average statistics, especially if they experienced a bad journey.
- *Big facts.* Big facts like the top causes of delay or the total number of journeys carried each year help customers put reliability performance and their own experience into context. But these big facts are not reliability metrics per se.
- *Future focused.* Customers welcome metrics that could help them plan their future journeys, as long as they translate into their personal experience. Such metrics include train frequencies, platform wait time and curiously, number of signal failures.

The big facts and future focused metrics have the most potential for further development. However, customers have their own perceptions of reliability, and most of them feel like they do not need any metric to prove what they experience on a daily basis. They argue that TfL needs to take action, and not publish metrics that are not meaningful to them and their experiences.

2.3 Previous Work on Journey Time Reliability Using Automated Data Collection Systems

Furth and Muller (2006) argue that the reliability measures currently used by most transportation agencies do not capture the impact of reliability on passengers. More specifically, the median waiting time for high frequency bus services does not account for what the authors call “potential” waiting time. A better measure for reliability, from the passenger’s perspective, would be the 95th percentile waiting time, interpreted as a measure of “budgeted waiting time”. The “potential” waiting time is simply the difference between the 95th percentile and the median waiting times.

2.3.1 Journey Time Reliability in the London Underground

Chan (2007) extended the measure proposed by Furth and Muller (2006) to the entire journey time, using the London Underground as an example. Using AFC data, she measured the compactness of the journey time distribution for a specific origin-destination pair by calculating a *journey time reliability factor*. The proposed metric was defined as the difference between the 95th percentile and the median journey time.

Uniman (2009) extended the research further. He introduced the concept of *reliability buffer time*, denoted RBT, defined as the “*amount of extra time that passengers must budget above the typical journey time in order to arrive on time at their destination with a specified level of certainty*” (Uniman et al., 2010). Similarly to the journey time reliability factor, the RBT is calculated as the difference between an upper percentile value N and a measure of the typical journey time, denoted by M .

The typical duration of the journey, M , is chosen as the median journey time, rather than the mean of the distribution, as the median is not sensitive to outliers. The upper percentile N represents, from the passengers’ perspective, the likelihood of arriving on time if the same journey is repeated every day. Figure 2-1 illustrates the likelihood of delays given the RBT upper percentile.

As shown in Figure 2-1, the 90th percentile implies a delay once every 10 journeys, or once every two weeks if the same trip is made every day, while the 95th percentile implies a late arrival once every twenty journeys, or once a month.

The choice of the upper bound also depends on the desired sensitivity of the metric

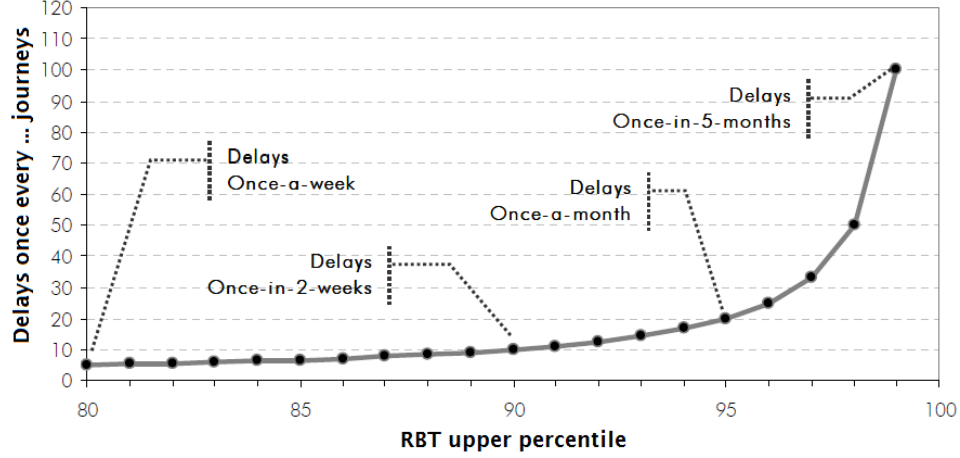


Figure 2-1: The likelihood of delays as a function of the RBT upper percentile value (adapted from Uniman, 2009)

to small delays. The higher the upper bound, the more sensitive the RBT is to small delays. The chosen percentile must also correspond to realistic objectives from an operational perspective.

Uniman (2009), as well as Furth and Muller (2006) and Chan (2007), suggest N to be the 95th percentile journey time. However, the transit agencies may choose different values based on their own operating characteristics.

The mathematical formulation of the reliability buffer time is given by:

$$\text{RBT}_{\text{OD}} = (TT_{95\%} - TT_{50\%})_{\text{OD}} \quad (2.2)$$

where $TT_{50\%}$ and $TT_{95\%}$ are respectively the median and the 95th percentile journey time, calculated at the O-D pair level, for a time interval that varies between 15 min to a full day, and over a period from 1 to 20 days.

The RBT for each O-D pair is then weighted by the passenger flows during the time interval of interest and an aggregate measure of performance is calculated for each Underground line, as follows:

$$\text{RBT}_{\text{line}} = \frac{\sum_{\text{OD} \in \text{line}} f_{\text{OD}} \cdot \text{RBT}_{\text{OD}}}{\sum_{\text{OD} \in \text{line}} f_{\text{OD}}} \quad (2.3)$$

where f_{OD} is the passenger flow on the O-D pairs of the line, and RBT_{OD} is the reliability buffer time calculated with Equation 2.2.

Uniman (2009) then extends the reliability buffer time metric with a classification of performance in two categories: *recurrent*, that reflects the inherent variability of journey time due to service characteristics, and *incident-related*, that captures the effect of service disruptions on the reliability experienced by the passengers. A reliability baseline is developed by classifying performance into these two categories.

Stepwise regression on the 95th percentile (referred to as the “*indicator of delay*”) is used to separate the recurrent and incident-related performance. For a set of observations, the proposed method answers the following question: “*If the i^{th} observation is removed from the sample, would the fit of the remaining observations around their mean be better than the fit of the previous set of observations (remaining + i^{th} observation) around their original mean?*”. The dependent variable Y is the 95th percentile journey time of each day for an O-D pair and time interval; the independent variables X_i are dummy variables such that $X_i = 1$ for the i^{th} observation, 0 otherwise. Figure 2-2 illustrates the comparison of the 95th percentile journey time across days for journeys made on a single O-D pair during a given time interval.

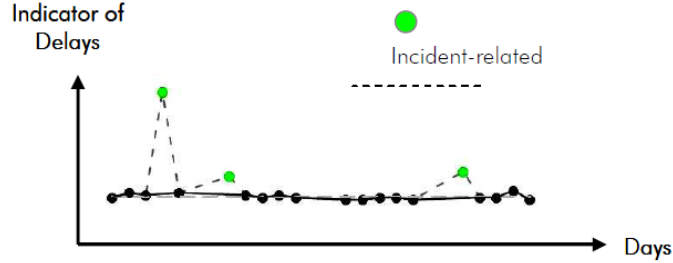


Figure 2-2: Classification into recurrent and incident-related performance
(from Uniman, 2009)

The recurrent and incident-related journey time distributions for each O-D pair and for the time interval of interest are then estimated by aggregating the journeys for each of the days classified in the corresponding category. Figure 2-3 illustrates the estimation of the recurrent performance, by pooling together all the days classified as recurrent.

The *excess reliability buffer time* (ERBT) is defined as “the amount of buffer time required by passengers to arrive on time with 95% certainty in addition to the amount of buffer time that would have been required under typical conditions” (Uniman et al., 2010), and is given by:

$$\text{ERBT}_{\text{OD}} = \text{RBT}_{\text{OD, overall}} - \text{RBT}_{\text{OD, typical}} \quad (2.4)$$

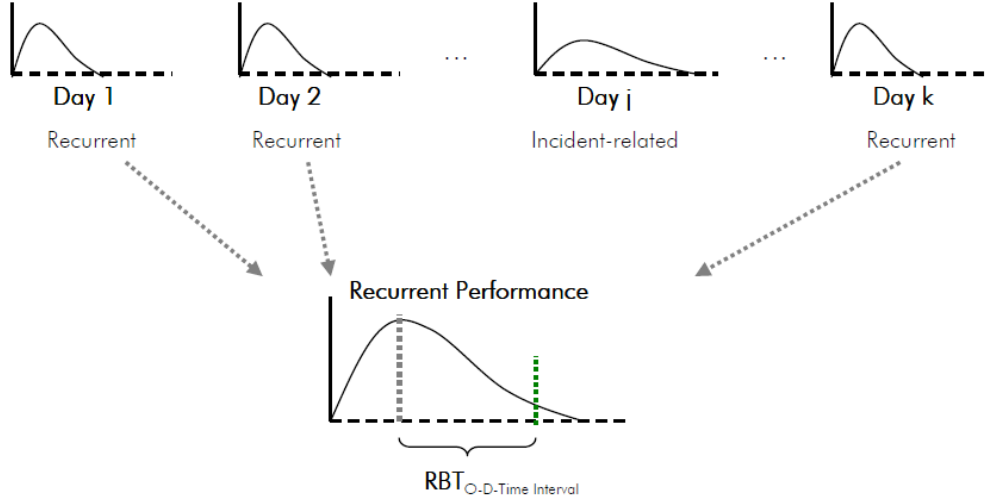


Figure 2-3: Definition of recurrent performance (from Uniman, 2009)

where $RBT_{OD, overall}$ is the actual buffer time experienced by passengers, which includes observations for all days, recurrent or incident-related, and $RBT_{OD, typical}$ represent the performance under recurrent conditions. Both are calculated for an O-D pair for a specific time interval over n days.

The excess buffer time can be calculated at the line level by weighting each O-D pair by its passenger flow:

$$ERBT_{line} = RBT_{line, overall} - RBT_{line, typical} \quad (2.5)$$

where $RBT_{line, typical}$ is the line level measure of the recurrent RBT.

Finally, Uniman (2009) defines a *percentage of unreliable journeys* metric that measures the likelihood of extreme delays. A journey is considered reliable if it is shorter than or equal to the 95th percentile journey time under typical conditions.

Percentage of Unreliable Journeys =

$$(\text{Percentage of Overall Journeys with J.T.} > RBT_{typical})_{OD} \quad (2.6)$$

Uniman (2009) applied all the proposed metrics to the London Underground, using Oyster card data, to evaluate the reliability of the system and validate the performance classification against the incident log.

The research showed the impact of incident-related performance and how disruptions affect reliability in a way that is not captured by traditional metrics that focus only on average performance. However, the classification based on stepwise linear regression is not simple to apply on a regular basis.

2.3.2 Reliability in the Bus Network

Ehrlich (2010) extended the work by Uniman (2009) with the definition of passenger-focused reliability metrics for the bus network in London, using data from London’s AVL system, iBus. Three measures of reliability are introduced: the *journey time* (JT), the *excess journey time* (EJT), and the *reliability buffer time* (RBT). JT includes the waiting time and the in-vehicle travel time. EJT is defined as the difference between the median journey time and the scheduled journey time. RBT is calculated following Uniman (2009) as the difference between the 95th percentile and the median journey time.

Ehrlich (2010) analyzed the reliability of the heaviest origin-destination pairs on several bus routes in London. His analysis extends London Buses’ current reliability metrics by considering bus journeys from the passenger perspective, and taking into account the entire journey experience, from the passenger’s arrival at the bus stop to his arrival at his destination. However, the work does not examine the extension of the metrics at the route level. Furthermore, it does not take into consideration the fact that O-D pairs, especially in busy corridors, may be served by more than one route. The existence of multiple routes serving the same O-D pair may reduce passenger waiting time. Additionally, the situation can be complicated if some of the routes are local, while others are express. In this case, passengers may develop strategies attempting to optimize the overall experience and consider the trade-offs between longer waiting times and faster travel times.

2.4 Origin-Destination Inference and Trip Linking

Gordon (2012) developed and implemented key utilities to process Oyster data and extract information used in the research presented in this thesis. The methodology infers bus trip origins and destinations and links trips into full, multimodal journeys. The algorithm uses four data sources:

1. *Automated Fare Collection (AFC)* (e.g. Oyster card data in London), that records the date, time, and location of each fare card transaction. In London, location includes entry and exit station in the Underground, and boarding transaction point for the bus system.
2. *Automatic Vehicle Location (AVL)* (e.g. iBus in London), that records the position of buses along the routes. Each record includes the vehicle's route, direction of travel, trip number, time of event, stop code and stop sequence number. The events of interest are vehicle arrivals and departures from stops.
3. *Station gateline counts* provide the total number of customers entering or exiting the station.
4. *Bus farebox counts* provide a record of each transaction, that includes time, vehicle trip number, route number and direction, and ticket type (e.g. Electronic Ticketing Machines (ETM) in London)

The main steps of Gordon's methodology are (Muhs, 2012):

1. *Origin inference.* For fully gated systems such as the London Underground, AFC records include the time and location of the customer's entrance into the system. However, the on-board fare collection system on London's buses does not record the location of the Oyster card validation. The algorithm infers the location of a bus customer's origin by matching the time stamp and vehicle trip number from the Oyster record to the time stamp and vehicle trip number in the iBus database.
2. *Destination inference.* Inferring destinations on a fully gated system such as the rail system in London, for which the user has to validate his Oyster card when exiting, is straightforward. However, for the bus network, the alighting location is inferred from the next smart card validation. A set of conditions is defined for the destination

inference. If the Oyster card bus transaction is the last of the day, it is assumed that the alighting location is the closest stop to the location of the first validation on the next day.

3. *Trip linking.* With the origin and destination inferences, Gordon (2012) also links trips into full journeys. Several conditions are set for which the trips are linked or not.
4. *Scaling.* The scaling process is needed since AFC transactions do not represent the entire population. Gateline counts and ETM box records provide an accurate estimate of the origin-destination flow for the entire transportation system, and are used to determine the scaling factors.

Muhs (2012) applies Gordon’s methodology to analyze changes in travel patterns after the opening of the East London Line Extension as part of the London Overground. In this thesis, the results from Gordon’s analysis are used to calculate passenger volumes on specific origin-destination pairs.

2.5 Conclusion

Traditional operation-focused reliability metrics, based on manual surveys and small sample sizes, only provide average values that (1) do not account for the entire passenger experience and (2) do not capture performance variability, by time of day, day of week, or season. Studies by Uniman (2009) and Ehrlich (2010) illustrate the use of AVL and AFC data in London to produce metrics that evaluate the whole passenger journey experience and capture performance variations. However, the two studies analyze the Underground and the bus network independently, and do not consider multimodal journeys. Moreover, the study by Ehrlich (2010) does not measure route level metrics and overestimates journey times on certain origin-destination pairs by ignoring overlapping routes. Finally, Uniman’s definition of recurrent performance does not provide a simple calculation of journey time reliability standards.

Chapter 3

Current Reliability Metrics used by Transport for London

This chapter presents an overview of the various reliability metrics currently calculated and published by the different operating subsidiaries of Transport for London. The chapter focuses on the London Underground and London Buses as the two major transportation modes of interest. The metric published by London Streets is also discussed in this chapter as it is the only measure that considers travel time distribution at a highly disaggregate level. The London Overground and the London Tramlink report an on-time performance metric, referred to as Public Performance Metric (PPM). This measure of timetable adherence calculates the percentage of trains that are less than 5 minutes late when they arrive at the destination terminal. The Docklands Light Railway also reports an on-time performance metric. These three modes will not be further discussed in this chapter as this research focuses on the Underground and the bus network.

Section 3.1 presents the performance metrics currently used by London Underground; section 3.2 discusses the reliability measures reported by London Buses, and section 3.3 briefly describes the metric published by London Streets.

3.1 London Underground

London Underground (LU) calculates and publishes various Key Performance Indicators (KPI) at the end of every four-week period. Some of the published metrics are targeted, i.e. the metrics for every period are compared to predefined targets that are updated every year.

Other metrics are not targeted and are just compared to the corresponding performance of previous periods. The targeted metrics include:

- *Customer Satisfaction.* The level of customer satisfaction is measured at the system level for every quarter with about 2,500 face-to-face surveys, in which customers are asked to score their overall experience as well as 19 train and station attributes.
- *Passenger Journeys.* The number of journeys made by fare-paying passengers is calculated every period for the whole Underground network.
- *Percentage of Schedule.* The actual kilometers operated are compared to the distance scheduled to be run and adjusted for planned short-term closures. This is a measure of the level of service from the operator's point of view. The percentage of schedule operated is split between peak and off-peak hours to measure the performance when the demand on the system is the highest in terms of passenger volumes and number of trains in service. The percentage of the timetable operated on weekdays is also compared to the performance on weekends to evaluate the impact of weekend engineering work on LU service.
- *Journey Time Metric.* The Journey Time Metric (JTM) is the customer-focused reliability metric currently reported by LU. The difference between the actual journey time and the scheduled journey time is referred to as the Excess Journey Time and reported as a targeted metric. The actual journey time is also reported by LU for customer information and referred to as Total Journey Time. The calculation of the JTM is detailed in section 3.1.1.

The non-targeted metrics include:

- *Station Closures.* Stations are fully closed when all entry and exits are closed and trains do not stop. The number of unplanned full station closures is calculated for every day of service and reported at the end of each period. Each station on the Underground is allocated to a specific line, and station closures are reported by line.
- *Escalator and Lift Availability.* The total hours escalators and lifts were working is compared to the scheduled service hours. Impact of planned maintenance and

unscheduled failures are taken into account. The escalator and lift performance is reported for each period by line, as for station closures.

- *Rolling Stock Mean Distance Between Failures.* This is a measure of fleet performance, from the operator’s perspective.
- *Lost Customer Hours.* Lost Customer Hours (LCH) reports disruptions and delays from the customer’s point of view. The calculation of the metric is detailed in section 3.1.2.

Table 3-1 summarizes the main Key Performance Indicators calculated and published¹ by LU, with the focus of the metric (customer or operator) and the level of aggregation.

Key Performance Indicators	Point of view		Aggregation	
	Customer	Operator	Line	System
<i>Targeted</i>				
Customer Satisfaction	x			x
Passenger Journeys	x			x
Percentage of Schedule		x	x	x
Excess Journey Time	x		x	x
<i>Non targeted</i>				
Total Journey Time	x			
Station Closures		x	x	x
Escalator and Lift Availability		x	x	x
Rolling Stock Mean Distance Between Failures		x	x	x
Lost Customer Hours	x		x	x

Table 3-1: Key Performance Indicators Evaluation

This section focuses on the Journey Time Metric as it is the main reliability metric currently used by London Underground; the non-targeted Lost Customer Hours metric is also discussed since it is closely related to the performance of the system in terms of reliability.

¹All the metrics are measured for each four-week period, except Customer Satisfaction which is measured quarterly.

3.1.1 Journey Time Metric

The Journey Time Metric (JTM) was defined in 1997 by the London Underground Market Planning department as a customer-focused reliability indicator (London Transport, 1999). JTM results are reported to the public at the end of every four-week period alongside the other Key Performance Indicators mentioned above.

The Journey Time Metric breaks the journey into four parts: Access, Egress and Interchange Time, Ticket Purchase Time, Platform Wait Time, and On-Train Time. In addition to these components, the JTM also considers the effect of line and station closures on customers' journey times. For each of the components, a value is estimated based on schedule, to reflect how long the journey would take if there were no disruptions. The JTM then compares the actual journey times to the scheduled ones. The difference between the two is referred to as Excess Journey Time (EJT). EJT is used as an indicator of the journey time reliability in the Underground.

Components of the Journey Time Metric

Access, Egress and Interchange (AEI) Time. The AEI time evaluates the walking time required to enter and exit the Tube and to transfer between two lines. The access time is measured from the station entrance at the street level to the mid-point of the platform; the interchange time is calculated between the mid-point of the arriving platform and the mid-point of the departing platform (within the same station); the exit time is the time from the mid-point of the platform to the station exit at the street level.

The AEI component of the JTM is calculated using manual survey timings taken at 27 stations that account for 46% of the Underground demand: the survey staff walks predefined routes at each station during the busiest weekday time periods. Several routes are surveyed at each station to account for their complexity and the multiple entries and exits; each station is visited at least 12 times each four-week period. The surveys are complemented by models that estimate the impact on AEI of delays and congestion due to lift and escalator failures, train service reliability, and demand variability.

The scheduled AEI time is calculated as the corresponding “free flow” walking times when stations are not congested.

Ticket Purchase Time (TPT). The TPT is the sum of the queuing time and the transaction time at the ticket office window or the automated ticket machines. The queuing time is surveyed on a regular basis; transaction times are recorded at all windows and for some ticket machines at the busiest stations. The scheduled TPT is calculated as 90% of the average transaction time from the previous year. All queuing time is considered excess journey time.

Platform Wait Time (PWT). The PWT is calculated as the time between the customer's arrival at the mid-point of the platform and the moment the boarded train departs. Customers are assumed to board the first train that serves their destination, or when there is no frequent direct service to their destination, they are assumed to board the first train and transfer at a convenient station from which frequent direct service to their destination is available.

The signaling system provides data on actual service headways for all the Underground lines except for the District line, for which the wait times are measured through manual surveys.

The average platform wait times are calculated for each section of the line and each time period for which the scheduled headways are supposed to be the same. The average actual platform wait time $\overline{\text{PWT}}$ is given by:

$$\overline{\text{PWT}} = \frac{\sum_i H_i^2}{2 \sum_i H_i} \quad (3.1)$$

where H_i are the headways recorded at every station of the line section during the time period of interest.

The scheduled passenger wait time at a station is calculated as half the scheduled headway for the corresponding section of line. Both calculations assume random passenger arrivals and constant passenger arrival rates.

An additional platform wait time, referred to as Left Behinds, is estimated as part of the JTM, to capture the extra time when passengers cannot board the first train because of crowding. The calculation of this additional time is based on demand levels and the regularity of train service.

On-Train Time (OTT). The OTT is calculated from the moment the train departs the origin station to the moment doors open at the destination station. The scheduled OTT is calculated from the operating timetable. The actual OTT is measured using data from the signaling system when available. For those lines where such data is not available (District, Piccadilly, and parts of the Metropolitan, Circle and Hammersmith lines), the OTT is obtained from a sample of signal cabin box sheets.

Closures. The Journey Time Metric takes into account the effect of closures and disruptions due to incidents and engineering work. Three types of closures are considered:

- *Unplanned short term closures and service disruptions.* These unpredictable closures are defined as disruptions that exceed 30 minutes. Their impact on customers is calculated after their occurrence.
- *Planned short term closures and service disruptions.* These closures are known in advance and advertised to customers. They last between 1 and 28 days, for instance weekend engineering works, and their impact on customers can be alleviated with alternative bus service. However, they inconvenience LU customers and are considered as a delay above the scheduled travel time.
- *Scheduled long term closures.* These are scheduled closures that exceed four weeks in duration, and they must be advertised to customers at least two weeks before the start date. The impact of scheduled closures is not reported as excess journey time, but rather is added to the scheduled journey time.

Value of Time

London Underground calculates both unweighted and weighted Journey Time Metrics. The weighted (or generalized) JTM was introduced to reflect the fact that customers perceive each component of their journeys differently. Value of Time (VOT) weights are attributed to the five components described in section 3.1.1, as summarized in Table 3-2.

Based on Table 3-2, the AEI time has different weights according to the layout of the station and the path that customers need to use to access the platform or transfer between lines. For instance, walking up stairs is considered as more onerous than riding escalators or walking through hallways, and is therefore given a weight of 4, while riding escalators

Travel Time Category	Weight
Walking (through ticket halls/passageways)	2
Walking up stairs	4
Walking down stairs	2.5
Combination of walking up and down stairs	3.25
Riding escalators/lifts	1.5
Ticket queuing time	3.4
Ticket purchase time	2.5
Waiting on platform	2
Left behind on platform	3
Traveling on train	variable

Table 3-2: VOT weights for JTM calculation

and walking through hallways have weights of 1.5 and 2 respectively. The value of time for the on-train travel time varies between 1 and 2.48, and depends on the average level of crowding in the trains for each period, calculated based on the demand and excess PWT results.

Figure 3-1 compares the contribution of each journey component to the total journey time for the unweighted and weighted metrics. On-train time constitutes 64% of the unweighted journey time metric, but only half of the generalized (weighted) JTM. This reflects the fact that although on-train time is the greatest part of the journey, other components weigh more heavily in passenger perception.

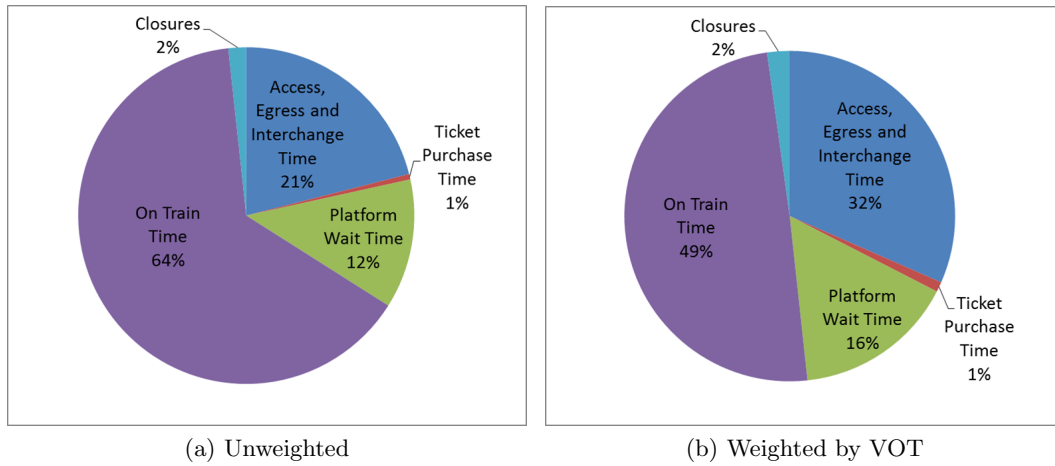


Figure 3-1: Contribution of each journey component to the Total Journey Time (using data for year 2011/12, excluding period 13)

Publication of the Journey Time Metric

The PWT and OTT components of the Journey Time Metric are estimated at the line section and timeband levels, and weighted by the passenger volumes obtained from the Rolling Origin-Destination Surveys to calculate average journey times at the line and period levels. The station related components (AEI and TPT) are calculated at the station level for each period. The performance of each station is then attributed to the “owning” line responsible for its management. The estimated impact of closures at the line level is also added to the journey time. The overall JTM at the network level is the average of the line performances, weighted by the line passenger flows (from the RODS data), and scaled up by a Journey Leg Factor (JLF) that accounts for the fact that about 40% of the London Underground customers interchange at least once during their journey, and therefore experience platform wait time and in-vehicle travel time more than once.

The line and network JTM results are included in the performance report published on the TfL website at the end of every four-week period. The performance of the latest period is compared to that of the previous ones and to the target set at the beginning of the year. Figure 3-2 shows the reliability metrics for the Jubilee line and for the whole network, as published at the end of period 9 (2011/12). Similar graphs are published for the other ten lines of the network.

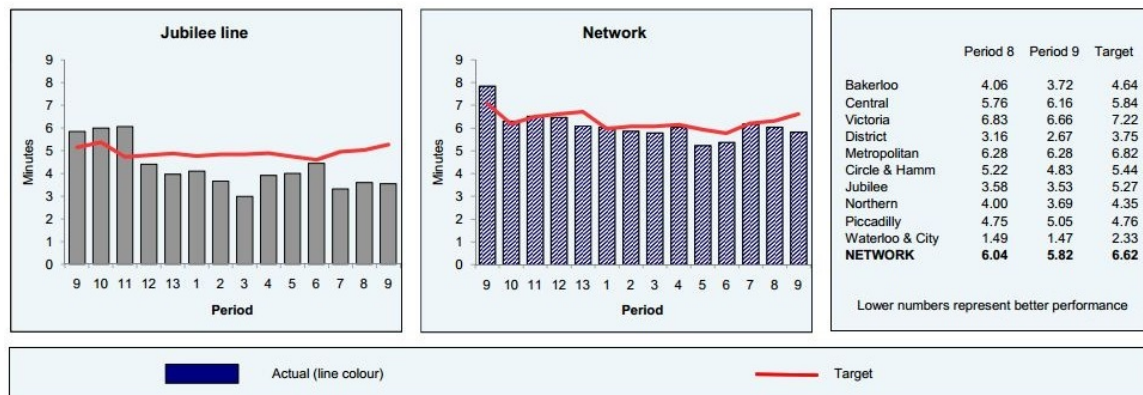


Figure 3-2: Weighted Excess Journey Time Results - Period 9, 2011/12
(from Transport for London, 2011b, p.7)

Internal performance reports break down the JTM results into the 5 components of the journey as defined in section 3.1.1, and analyze the key contributors (fleet, signals, staff, etc...) to the platform wait and on-train excess journey time. These results are used to

improve the overall performance of the Underground by taking actions targeted at the most deficient components of the journey experience.

Strengths and Weaknesses of the JTM Metric

Chan (2007) and Uniman (2009) have analyzed the strengths and weaknesses of the Journey Time Metric, which are summarized below:

Strengths of the current metric.

- The Journey Time Metric evaluates reliability performance from the passenger's perspective, by taking into account the entire journey experience, from the entrance at the origin station to the exit at the destination. Moreover, the value of time assigned to each component of the journey reflects customers' perceptions of each stage of their journey.
- The JTM breaks the journey into 5 components, which makes it a powerful operational metric as actions can be targeted in response to poor performance on specific components of the journey. The performance can also be attributed to specific lines and stations.

Weaknesses of the current metric.

- Although the calculation of the Platform Wait Time and the On-Train Time is mostly based on automatically collected data, the Access, Egress and Interchange time and the Ticket Purchase Time components of the JTM rely heavily on surveys and models, leading to costly data collection, small sample sizes, and infrequent updates of these components.
- The JTM presents average values over a four-week time period and at the line or network levels. It does not provide detailed information about travel experience at a more disaggregate level. The JTM, for example, cannot be estimated at individual journey levels, as the data used and the collection methods do not provide information for specific origin-destination pairs.

- Even within the same time interval, passengers' experience can vary because of service variability and passenger behavior. The JTM does not measure variations between journeys as it is simply an average metric.
- The performance of each station, and therefore the AEI and TPT components of the journey time, are attributed to a single line. This is intended to simplify the management of the station by giving full responsibility of each station to a single line manager. However, this can bias the performance results, especially for the lines responsible for the major interchange stations in Central London. Chan (2007) cites the example of the Oxford Circus, the most heavily used station in the network, which is served by the Victoria, Central and Bakerloo lines, but is under Bakerloo line management.

3.1.2 Lost Customer Hours

The Lost Customer Hours (LCH) metric is not a measure of reliability per se, but rather a measure of the availability of service. However, it is a very strong indicator of service reliability and is a reference metric at TfL. The LCH will be compared to the proposed set of metrics in subsequent chapters.

The LCH metric estimates the total impact on customer time of any delay or disruption lasting more than two minutes, considering the duration, location and time of day of the disruption. For example, a five-minute delay in the morning peak at Oxford Circus will have a LCH cost that is significantly greater than that of a delay of the same duration that occurred in the suburbs on a Sunday morning (Transport for London, 2011b).

Lost Customer Hours measures the impact of disruptions and translates the fact that an incident might force customers to choose another path or mode to complete their journeys. The goal for London Underground is to minimize LCH.

Figure 3-3 shows the Lost Customer Hours metric as published on the TfL website at the end of every period. LCH for the Jubilee line exhibits significant variability from one period to another. LCH for the network as a whole is more stable, especially during the last ten periods.

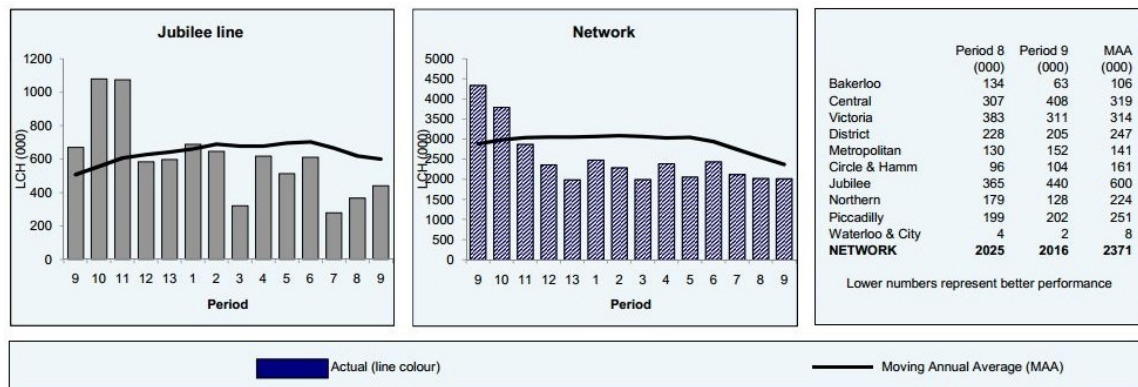


Figure 3-3: Lost Customer Hours - Period 9, 2011/12
(from Transport for London, 2011b, p.18)

3.2 London Buses

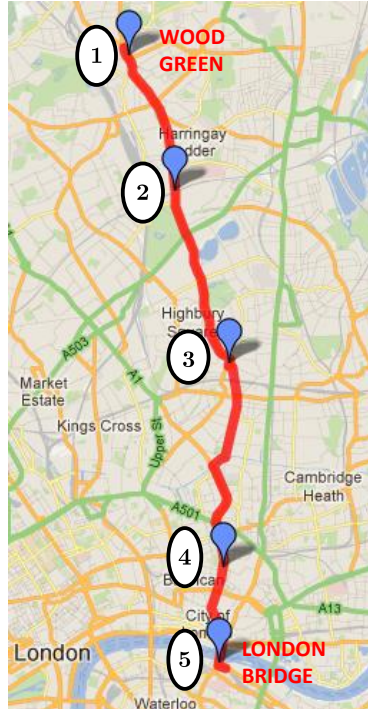
This section describes the service reliability metrics that London Buses currently uses to monitor performance. London Buses measures reliability through its Quality of Service Indicators (QSI). QSIs are calculated from data collected by manual surveys². These surveys are undertaken at about 500 locations on the bus network, called QSI points. Each of these QSI points is surveyed 16 times per quarter, with each observation lasting up to three hours. Bus departure times at the QSI points are recorded in order to calculate the headways and estimate customer waiting times. Figure 3-4 shows the QSI points for route 141, a high frequency route that operates between London Bridge and Tottenhall Road in the borough of Enfield (north London).

There are a total of five QSI points for route 141; two are surveyed in the northbound direction only, one in the southbound direction only, and the remaining two are surveyed in both directions. The “Shifts” represent the time periods during which each QSI point is surveyed. There are 8 different shifts in a week: shifts 1 to 5 are on weekdays and shifts 6 to 8 for weekends. For instance, the London Bridge QSI point is surveyed during all shifts, while the Newington Green northbound QSI point is surveyed only during weekdays and Saturday mornings.

For high frequency routes³, the Quality of Service Indicators include:

²QSIs will soon be calculated from iBus data, which will provide almost a 100% sample of headways and running times.

³High frequency routes have five or more trips per hour; passengers tend not to consult the schedule and are assumed to arrive randomly. 70% of the bus routes in London are high frequency routes.



	Observation Points	Shifts	Directions
1	Wood Green High Road	1-8	SB
2	Harringay – Green Lanes	1-8	Both
3	Newington Green	1,2,6	SB
	Newington Green	3,4,7,8	NB
4	Moorgate – Finsbury Sq.	1-8	NB
5	London Bridge III	1-8	NB

Figure 3-4: Route 141 - Map and QSI points

- *Excess Wait Time*
- *Percentage chance of waiting* less than 10 minutes, 10-20 minutes, 20-30 minutes and more than 30 minutes.

These two indicators are described in sections 3.2.1 and 3.2.2 respectively.

For low frequency routes⁴, the focus is on schedule adherence, as many customers rely on the schedule to plan their arrival time at the stop.

For each route and QSI point, the following indicators are reported:

- *Percent On-Time.* A bus is considered on time if it departs the QSI point between 2.5 minutes early and 5 minutes late.
- *Percentage chance of a bus running early.* A bus is early if it departs more than 2.5 minutes before the advertised time.
- *Percentage chance of a bus running late.* A bus is late if it runs between 5 and 15 minutes late.

⁴Low frequency routes have four or less trips per hour; passengers are more likely to use the schedule to minimize their waiting time and punctuality is critical for these routes.

- *Percentage chance of a bus not running.* This is the probability that a bus is more than 15 minutes late or does not run at all.

In addition to the QSI, London Buses also reports for all bus routes the vehicle kilometers operated and the percentage of kilometers lost for staff, mechanical or traffic reasons.

This section focuses on the calculation of the Excess Wait Time as it is the main performance metric reported by London Buses for high frequency routes. The percentage chance of waiting more than a certain threshold is also discussed.

3.2.1 Excess Wait Time

For high frequency bus routes, the Excess Wait Time (EWT) is calculated as the difference between the actual and scheduled wait times. This measure of performance is calculated at the QSI points. An average Excess Wait Time is then calculated for each route and operator, for each quarter.

Calculation of the Metric

Equation 3.2 gives the general formula for the calculation of EWT:

$$\text{EWT} = \text{AWT} - \text{SWT} \quad (3.2)$$

where the Average Waiting Time (AWT) and the Scheduled Waiting Time (SWT) are given by Equations 3.3 and 3.4 respectively.

$$\text{AWT} = \frac{\sum_i H_i^2}{2 \sum_i H_i} \quad (3.3)$$

$$\text{SWT} = \frac{\sum_i (H_i^S)^2}{2 \sum_i H_i^S} \quad (3.4)$$

where H_i are the actual headways measured from the data collected manually at a QSI point on the route during the time period of interest and H_i^S the corresponding scheduled headways.

The Schedule, Actual and Excess Waiting Times are calculated for every route, direction, QSI point, and shift. Weighted metrics are then calculated at the QSI point, shift, route

and direction levels: AWT is weighted by observed buses per hour (OBPH) and SWT is weighted by scheduled buses per hour (SBPH) (London Buses, 2002).

The aggregate EWT is then calculated at the route, operator or network level for each quarter as follows:

$$\text{EWT}_{\text{aggregate}} = \frac{\sum \text{AWT} \times \text{OBPH}}{\sum \text{OBPH}} - \frac{\sum \text{SWT} \times \text{SBPH}}{\sum \text{SBPH}} \quad (3.5)$$

The sums in Equation 3.5 depend on the level of aggregation required. For the route level, the sums are calculated over all the QSI points and shifts surveyed during the quarter, for both directions. For the operator or the system level, the summation is calculated over all the routes operated by the same contractor or all the routes in the system. The metrics are produced for each quarter and published on the TfL website.

Strengths and Weaknesses of the Metric

Waiting time at a bus stop is usually perceived by the customers as the most bothersome part of the journey experience. The use of the Excess Wait Time metric as the main Quality of Service Indicator emphasizes TfL's desire to measure and improve this component of the journey experience. However, EWT does not account for the entire journey experience, from the moment a customer arrives at the bus stop to the moment he alights at his destination stop. The in-vehicle travel time, even though often perceived as less bothersome than the wait time, is the longest part of the journey, and therefore should be accounted for when assessing performance of the bus network.

Additionally, EWT is estimated at only a few QSI points along each route. Manual surveys are also required to calculate the metric. Although the iBus system provides information that allows the calculation of EWT at all stops and for all times of day, this data is not currently utilized. This will change in the short term, as London Buses is planning to use iBus data to measure its QSIs.

Finally, EWT is mostly an operator-focused metric. The metric is calculated only at the route level, while customers are interested in their complete journey, from their origin to their destination. EWT also does not account for the fact that passengers might have several alternative routes to reach their destination, when multiple overlapping routes serve the same origin-destination pairs.

3.2.2 Percentage Chance of Waiting More Than a Given Threshold

London Buses reports for each route, each operator, and the whole network a set of four metrics that give an indication of waiting time variability. These metrics show the probability of waiting less than 10 minutes, 10 to 20 minutes, 20 to 30 minutes, and more than 30 minutes. The probabilities are calculated from the headway data taken at the QSI points of each bus route.

These measures are purely informational and are not compared with target values. These metrics are relatively simple and understandable to the London Buses customers, even more so than the Excess Wait Time metric. They are customer-focused as they try to capture individual waiting times. However, the calculation of the metric is based on the manual surveys and the small sample sizes. The percentages do not account for the entire population and do not reflect the variability of the headway distribution.

3.2.3 Publication of the Metrics

The metrics are published at the end of every quarter, for each route, operator, and at the network level.

Table 3-3, using data from the latest quarterly report produced and published by London Buses (Transport for London, 2012b), compares the reliability metrics for the high frequency bus routes for the third quarter of 2011/12 with the results for the same quarter of 2010/11. The numbers show a small overall improvement in the reliability of the bus network.

High Frequency Services	Third Quarter 2011/12	Third Quarter 2010/11
Average scheduled wait (minutes)	4.39	4.41
Average excess wait (minutes)	1.07	1.18
Average actual wait (minutes)	5.46	5.58
% Change of waiting <10 minutes	85.7%	84.7%
% Change of waiting 10-20 minutes	13.1%	13.8%
% Change of waiting 20-30 minutes	1.1%	1.4%
% Change of waiting >30 minutes	0.2%	0.3%

Table 3-3: London Buses Network Performance (from Transport for London, 2012b)

Figure 3-5 illustrates the long term trend in Excess Wait Time over the past twenty years. The graph shows a significant decrease of the EWT around 2003. This decrease may be attributed to the introduction of the London congestion charge in February 2003,

which reduced the number of cars in Central London, thus improving traffic conditions. Between 2004/05 and 2009/10, the EWT remained stable, before slightly decreasing in 2010/11. Results for 2011/12 will confirm if 2010/11 was the beginning of a further long term decrease or if it was just a particularly good year.

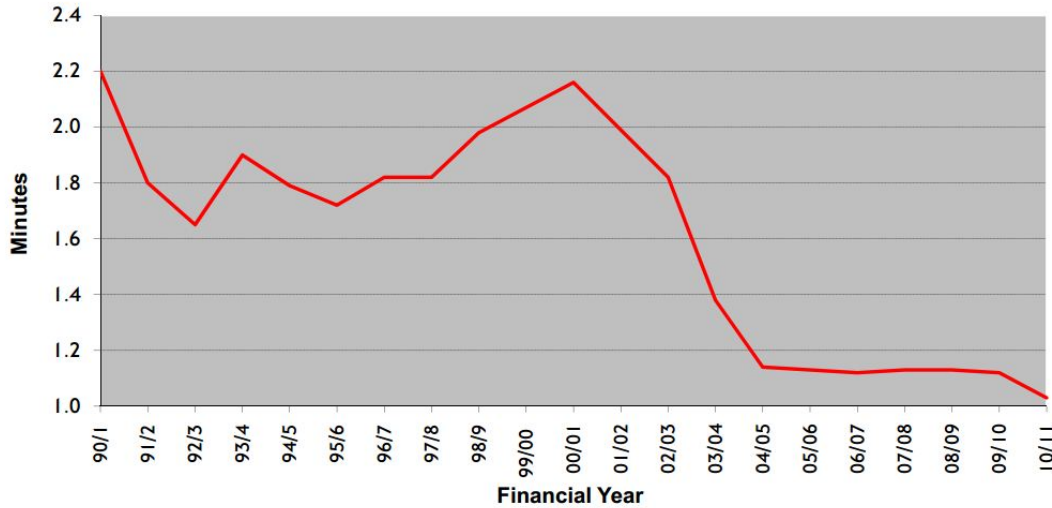


Figure 3-5: Excess Wait Time for high frequency bus services - Long term trend (www.tfl.gov.uk/assets/downloads/businessandpartners/long-term-trends.pdf)

3.3 London Streets

London Streets is the authority in charge of road and traffic management in Greater London. London Streets manages and operates the Transport for London Road Network (TLRN), 580 km of major roads that represent 5% of the London road network but carry more than 30% of the capital’s traffic. London Streets also manages the Congestion Charging scheme.

London Streets calculates and reports its own metric for travel time reliability, the Journey Time Reliability (JTR). In contrast to the metrics used by the London Underground and London Buses, described in the previous sections, the JTR looks at the travel time distribution for 5-minute time intervals on select links of the TLRN. The metric is calculated at a very fine spatial and temporal resolution, and then aggregated for corridors or the full network (each timeband separately).

JTR is defined as “the percentage of nominal 30 minute average length journeys⁵ com-

⁵30 minutes represents the average journey time of a typical commuter travelling by car across London (Emmonds and Turner, 2010).

pleted within 35 minutes” (Emmonds and Turner, 2010). The JRT is calculated with data from the Automatic Number Plate Recognition (ANPR) system (using cameras located at selected intersections). A link is defined as a section of road between two ANPR cameras. For each link, journey times are recorded and the mean journey time is calculated for each 5-minute interval within the AM or PM peak periods. An acceptable threshold is calculated for each link, each week and each 5-minute interval within the peak hours. The threshold is normalized so that it is equivalent to that for the average 30 minute journey time: the acceptable travel time is the average weekly travel time on the link and during the 5-minute interval of interest, multiplied by 1.167 or a ratio of $35/30$. Each daily mean is compared to the threshold for acceptable journey time, and based on this comparison, the 5-minute interval of interest is considered “acceptable” or “unacceptable”. Figure 3-6 illustrates this methodology for a single link and a single 5-minute interval.

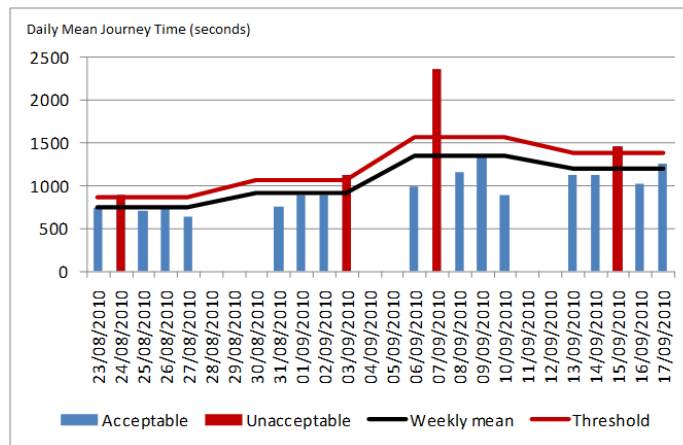


Figure 3-6: Example of determining which journey times are acceptable / unacceptable for a single link for the 7:00-7:05am interval (from Emmonds and Turner, 2010)

The percentage of acceptable 5-minute intervals in the AM and PM peak periods is calculated for each quarter and for each link. The results are then aggregated at the corridor level, using vehicle flows and link lengths as weights. An overall JTR is also calculated for the entire TLRN.

Figure 3-7 shows the JTR values for the entire Transport for London Road Network and for Central London, and compares the four quarters over the last 3 financial years. These graphs are published on the TfL website and updated every three months.

The Journey Time Reliability metric calculated by London Streets is the most disaggregate of the metrics published by Transport for London. It is also the only metric that

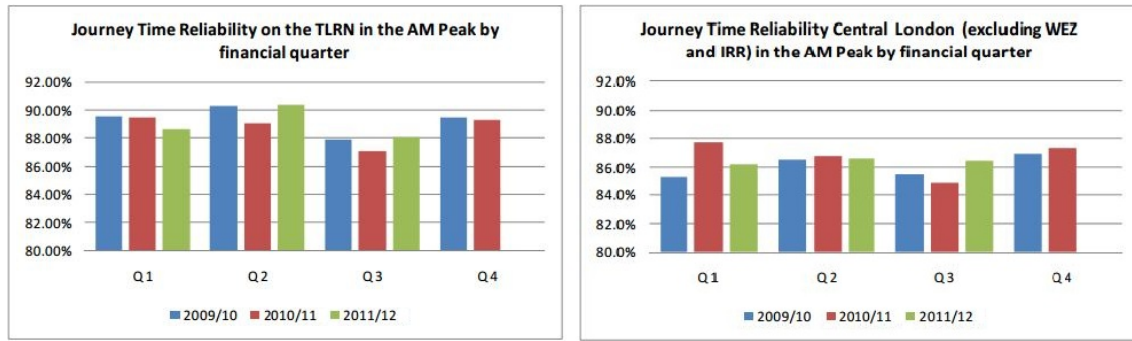


Figure 3-7: London Streets Journey Time Reliability - Quarter 3 2011/12
(from Transport for London, 2012c, p. 3)

accounts for the inherent variability of journey times, even under normal conditions. Even though the published values are aggregated at the corridor or network level, and for an entire quarter, the calculation methodology allows an analysis of the results at more disaggregate spatial and temporal levels. The methodology used by London Streets for the calculation of JTR provides a model for the performance metrics that could be developed for the other modes if similar journey time data was used for the estimation of the journey time reliability for origin-destination pairs and short time intervals.

3.4 Conclusions on the Current Reliability Metrics

The reliability metrics reported to the public by both London Underground and London Buses only show average values over a four-week period or a quarter. Averages do not measure the variation of service performance by time of day, day of week, or season. Additionally, the metrics are calculated at the line or route level and do not provide values at the origin-destination pair level. Customers do not have information for their own journeys. Finally, the Excess Waiting Time reported by London Buses does not consider the entire journey experience while the Journey Time Metric for the Underground does. The Journey Time Reliability, reported by London Streets, is calculated at a fine spatial and temporal resolution, and provides a model for the other modes.

Adequate metrics for journey time reliability should take into account the entire journey experience and the variability of journey time; they should also be similar for all the modes in order to provide reliability measures for multimodal journey and for comparison across modes.

Chapter 4

Measuring Journey Time Reliability in the London Underground

A London Underground customer is planning a trip from Canary Wharf to Bond Street, as shown on the map in Figure 4-1. For this purpose, he looks at the journey planner on the Transport for London website¹. Figure 4-2 shows the result of his request, for a trip during the PM peak on Monday, November 14, 2011. The journey planner gives an average and a maximum journey time of 16 minutes. From a trip planning point of view, this implies very reliable service. However, the actual situation may be quite different. Figure 4-3a presents the journey time distribution derived from Oyster card data for the same journey, in the PM peak for weekdays between November 14 and November 18, 2011. For Monday, November 14, the distribution spans from 18 to 38 minutes.

Oyster data records journey times from the origin station fare gate to the destination station fare gate while the journey planner travel time is measured from platform to platform, which explains why none of the recorded Oyster journeys were made in 16 minutes. However, the journey planner does not report the variability of journey time and provides an average and a maximum journey time that are identical (at least in this case).

Figures 4-3b and 4-3c show the travel time distributions for two additional origin-destination (O-D) pairs for the same week in November, and for the PM peak. The three

¹<http://journeyplanner.tfl.gov.uk>

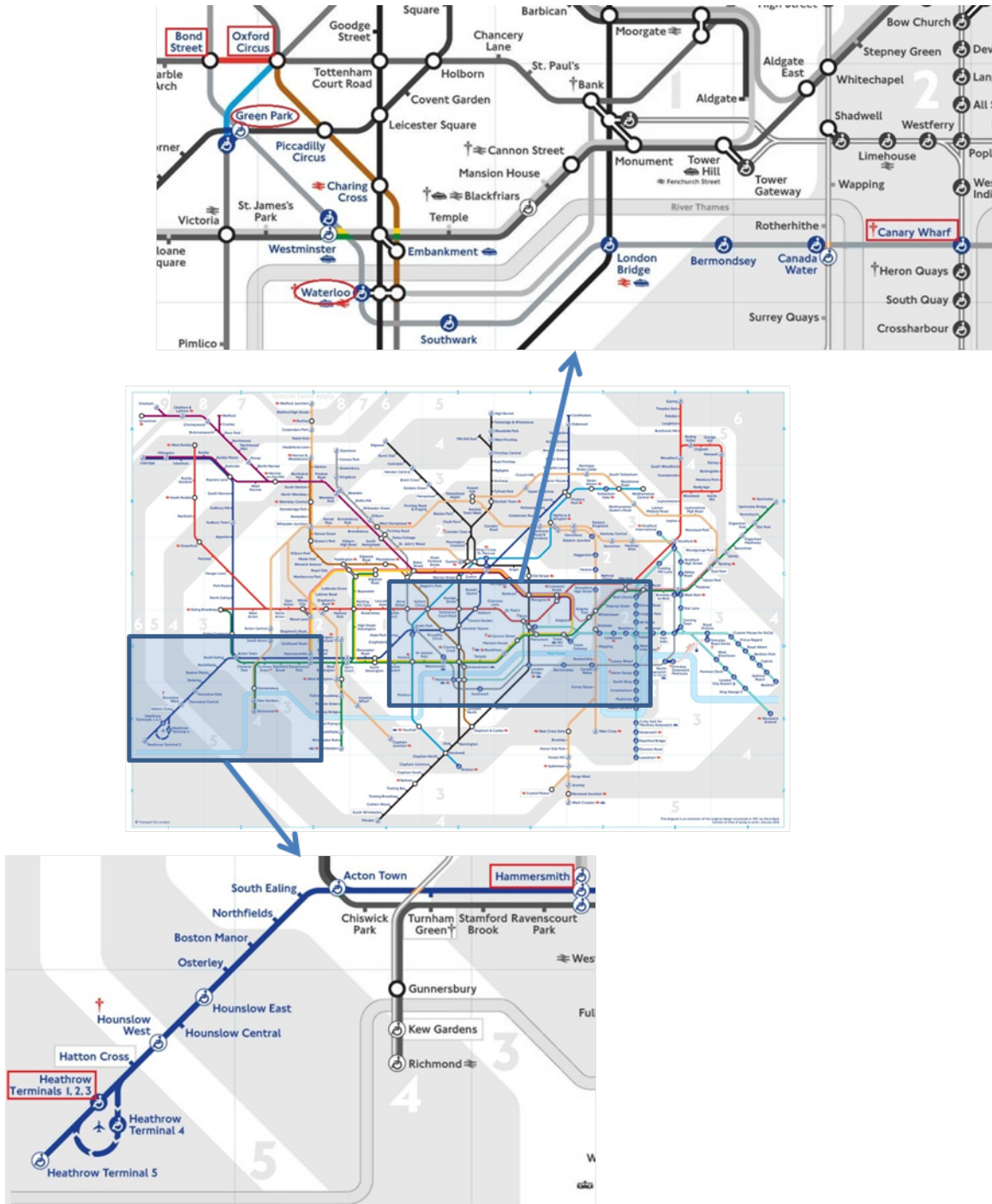


Figure 4-1: Map of the 3 origin-destination pairs: Canary Wharf to Bond Street, Heathrow (Terminals 1,2,3) to Hammersmith, Canary Wharf to Oxford Circus

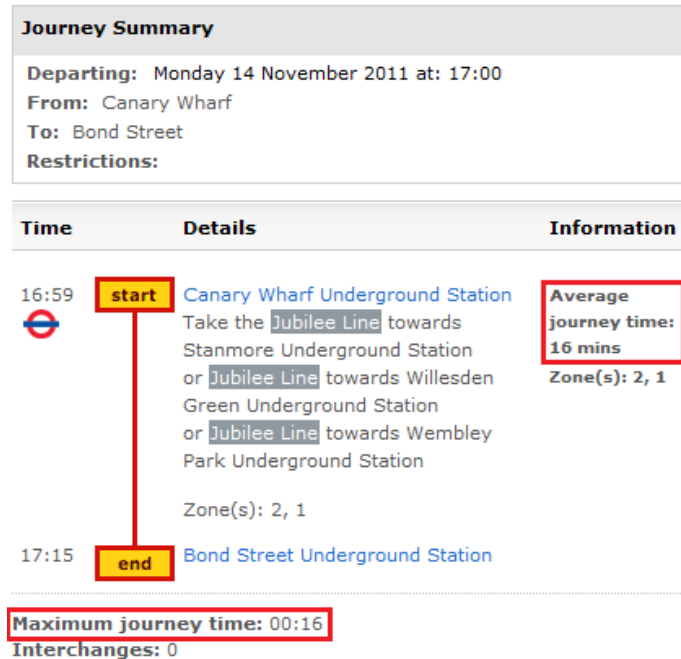
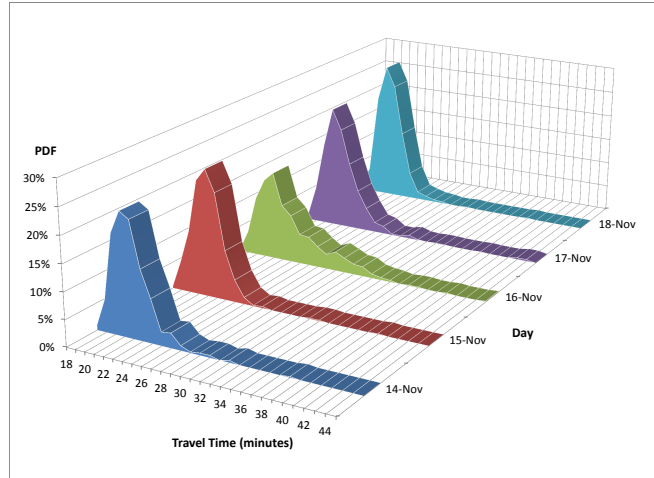


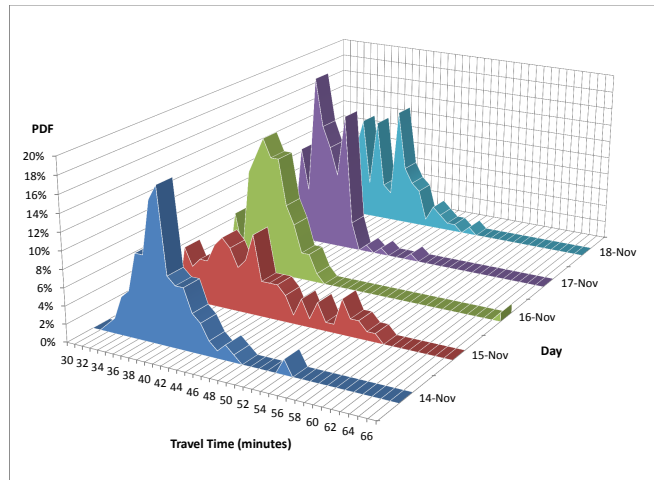
Figure 4-2: Journey planner - Canary Wharf to Bond Street

pairs, presented on Figure 4-1, are chosen to represent a variety of trips that can be made in the Underground network. The journey from Canary Wharf to Bond Street is a direct trip in Central London on the Jubilee line. The Heathrow (Terminals 1,2,3) to Hammersmith journey on the Piccadilly line is direct but longer, starting in zone 6 and ending in zone 2. The journey from Canary Wharf to Oxford Circus, also in Central London, involves an interchange and a path choice: there is no direct service available between these two stations. Customers can use the Jubilee line and then transfer either at Green Park (to the Victoria line), at Bond Street (to the Central line), or at Waterloo (to the Bakerloo line).

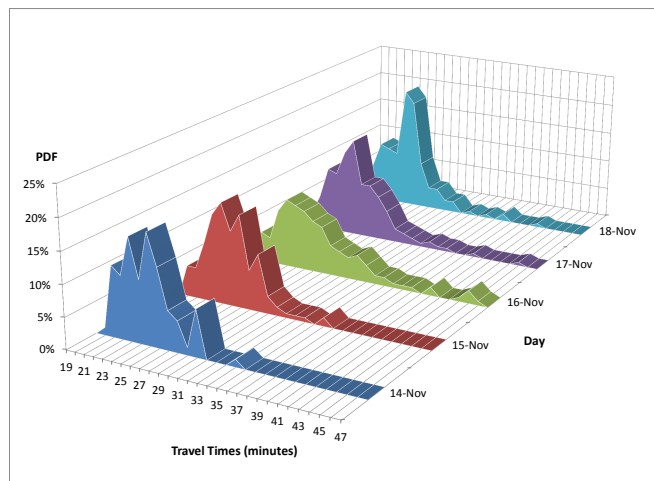
The figures for the three origin-destination pairs illustrate the journey time distributions and the corresponding variability even within the same day. The distributions for the Jubilee line journeys from Canary Wharf to Bond Street are narrow, except for November 14 where the longer journeys reflect the minor delays on the line due to an incident that occurred between 17:00 and 18:00. The larger spread of the distributions for the journeys from Heathrow to Hammersmith can be explained in part by the average 5-minute headway in the PM peak, when there is a train every 2 to 3 minutes on the Jubilee line in the same time period. The spread of the distributions for the journeys between Canary Wharf and Oxford Circus can be explained in part by the interchange that introduces a variability



(a) Canary Wharf to Bond Street - Jubilee line



(b) Heathrow Terminals 1,2,3 to Hammersmith - Piccadilly line



(c) Canary Wharf to Oxford Circus

Figure 4-3: Journey time distributions for 3 O-D pairs - PM peak

that depends on the passengers' individual behavior and preferences (path choice, walking speed, ...). The variability of journey time is also based on the performance of two Tube lines.

These three origin-destination pairs illustrate some of the challenges that arise when studying journey time reliability for the London Underground, such as the interchange and path choice issues that need to be accounted for when measuring journey time reliability in the LU network.

This chapter proposes a methodology to measure journey time reliability in the London Underground for weekdays², that takes into account various levels of spatial and temporal aggregation. The methodology is based on the comparison of the travel time distribution on one specific day with the distribution on other days, for the same O-D pair and at the timeband level³. An indicator of travel speed and a measure of the spread of the travel time will be provided for each origin-destination pair and each line in the system.

This chapter is organized as follows. Section 4.1 describes the requirements for a good journey time reliability metric. A framework for journey time reliability measurement is proposed in section 4.2, with an analysis at the origin-destination pair level, and the aggregation of the metrics at the line level. Section 4.3 proposes an additional metric for comparison of Underground lines. Results of the metrics previously defined are shown in section 4.4. The use of the metrics is discussed in section 4.5. Section 4.6 gives a brief assessment of the proposed metrics.

²The weekends are not considered in this thesis. Transport for London is currently undertaking an update of the Tube network. This improvement plan involves partial or total closures of some lines during weekends for engineering works. These closures force customers to adjust their travel plans, and the journey time increases do not reflect any unplanned disruptions. Additionally, if a section or an entire tube line is closed, no reliability performance can be calculated for the corresponding weekends.

³ Each weekday is decomposed into three time periods, or timebands, defined as follows:

- AM peak from 7:00 to 10:00
- PM peak from 16:00 to 19:00
- Off-peak for the rest of the weekdays.

4.1 Requirements for the Journey Time Reliability Metrics

This section describes the requirements for an appropriate reliability metric. When setting requirements for a reliability, one must keep in mind the two different points of view defined in Chapter 1. The operator and the customers don't have the same criteria for an adequate metric, and a good measure should fulfill the requirements from both sides. The criteria proposed here and derived from Chan (2007) and Uniman (2009) consider these two perspectives.

- *Customer-driven.* The metric must capture passengers' experience and perception of the journey time reliability. The abundant Oyster dataset provides information on the full range of service performance experienced by the customers and allows the definition of metrics that take into account the entire journey experience and the variability of journey time instead of calculating only an average.
- *Simple.* The metric must be simple to compute in order to allow regular publication of the results. The metric must also be easily understandable so that it requires little explanation when published for internal use at Transport for London or presented to the public.
- *Meaningful for the customers.* The metric must be useful for customers, give them insights on the performance of the Underground system and more specifically for their own journey. The metric should also help passengers plan their future trips.
- *Meaningful for the operator.* The simplicity of the metric must not compromise its utility for the operator. A good measure for journey time reliability must translate into practical actions to improve the performance of each line, and reciprocally the metric must reflect the actions taken to improve the performance.
- *Standardizable.* The metric must allow the definition of journey time reliability standards to compare the daily performance to the expectations in terms of journey times and reliability.

4.2 Reliability Measurement Framework for a Single Underground Line

Traveling in the London Underground with an Oyster card requires entry and exit validations in order to be charged the appropriate amount for the trip. For each journey, this validation sequence provides the exact travel time between the origin station and the destination station. Assuming that several passengers travel on the same origin-destination pair, a journey time distribution can be obtained for every O-D pair traveled in the network.

This section provides a framework building on previous work by Chan (2007) and Uniman (2009) for travel time reliability measurement at the line level. The proposed methodology is decomposed into three main steps.

1. Analysis at the origin-destination pair level. This is the most disaggregate level of aggregation and also the most relevant for customers.
2. Aggregation at the line level, to produce metrics that are more relevant for the operator.
3. Definition of reliability standard and additional metrics.

These three steps will be detailed in this section.

4.2.1 Data Available

The data used for this analysis consists of the Oyster records for trips made in the Underground, from one station to another. The focus of this thesis is on the Tube but the methodology can be applied to any other rail mode in London as long as entry and exit validations of the Oyster card are required.

Error in the Oyster Data

The Oyster data records the transaction time truncated to the minute⁴. This means for instance that the system will record 17:00 for all the transactions that occurred between 17:00:00 and 17:00:59. As pointed out by Chan (2007), this truncation of the transaction

⁴TfL is currently upgrading their system and will soon be able to record transaction times without truncating the seconds.

time introduces an error of ± 59 seconds since both the entry and exit validations are recorded at the minute level. Figure 4-4 illustrates this margin of error for a single journey.

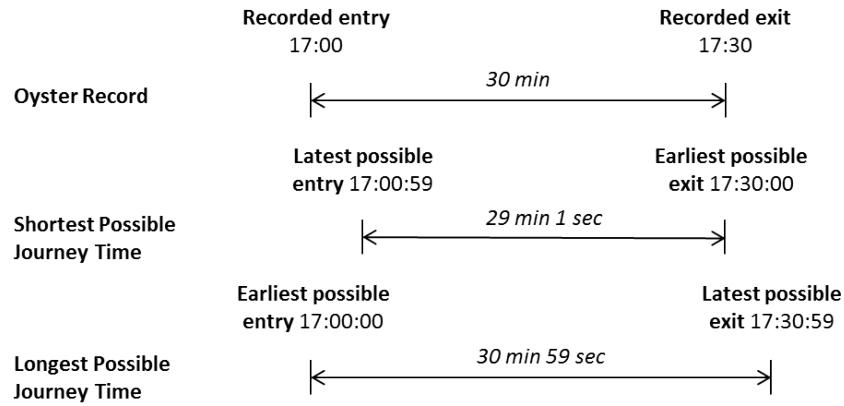


Figure 4-4: Margin of error of the Oyster journey time for an individual journey - Adapted from Chan (2007) and Uniman (2009)

Another way to look at the problem considers two passengers who travel together in the Underground. Even if these two customers travel together, the exact times when they validate their Oyster cards at the entry and the exit station will likely differ by a few seconds. One can easily imagine that one passenger validates his card 1 second after the other. The extreme case, illustrated in figure 4-5, shows one passenger entering the system at 17:00:59 and exiting at 17:30:00, when the second customer enters at 17:01:00 and exits at 17:29:59. This would result in a recorded Oyster travel time of 30 minutes for the first passenger and 28 minutes for the second, and thus a 2-minute difference in the Oyster travel time when the actual difference is 2 seconds.

The margin of error as described in the two examples above can readily explain a 2-minute variation of the Oyster travel time. Recording the Oyster validation times without truncating the seconds would increase the precision of the journey times and thus diminish the margin of error.

4.2.2 Origin-Destination Pair and Timeband Analysis

As described earlier in this section, Oyster records show that the journey times from one Underground station to another are not constant, but rather are distributed with a span that reflects the travel conditions on the Underground lines of interest, during the time period that is analyzed.

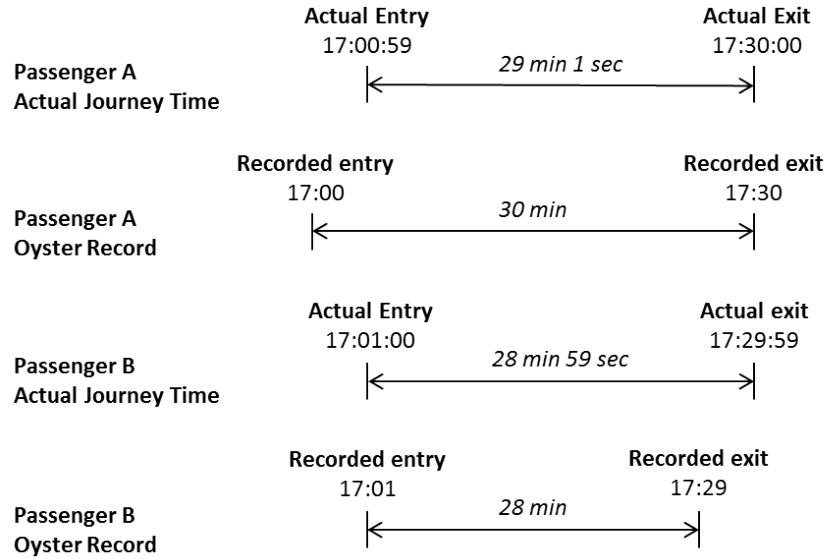


Figure 4-5: Margin of error of the Oyster journey time - Simultaneous trips

Journey Time Distribution

The journey times for passengers traveling between two stations on the same day and during the same timeband most likely vary. As a consequence, a journey time distribution for any origin-destination pair traveled can be obtained with the Oyster card data. The central tendency and the spread of these distributions are critical. The central tendency is an indicator of what the “average” customer can expect for his journey. Ideally, the spread should be as small as possible since all passengers should make the trip in the same time in a highly reliable system.

The origin-destination pair and the timeband are the finest spatial and temporal levels at which a journey time distribution can be developed. For every timeband and for every origin destination pair traveled in the Underground, a journey time distribution can be developed for this analysis. Additionally, the customers have interest in the reliability performance of their own journeys. Because of the customer’s interest in making specific journeys, it is relevant to do the analysis at the O-D pair and timeband level. The metrics can then be aggregated appropriately to reflect the operator’s interest.

Uniman (2009) proposed an analysis of the journey time distribution at the O-D pair level for a time period ranging from 15 minutes up to an entire timeband, using a sample size of 1 to 20 weekdays. This thesis looks at the journey time distribution at the timeband

level and considers each weekday independently. The distributions for the same O-D pair and the same timeband are compared across days. An ultimate goal for Transport for London would be to have a similar journey time distribution every day, for each origin-destination pair in the Underground network.

This thesis proposes the timeband as the most disaggregate temporal level of analysis. The morning and evening peaks should be studied independently for operational purposes and to be consistent with the Key Performance Indicators (KPI) currently published by Transport for London. The peaks are considered as the critical timebands for the Underground; it is therefore valuable for TfL to see metrics at the timeband level, so that peak period performance is distinguished from off-peak performance.

Reliability Buffer Time

The reliability buffer time (RBT), is defined in section 2.3.1 as the difference between the 95th percentile and the median of the journey time distribution. Equation 4.1 presents the mathematical formulation of the reliability buffer time for a specific origin-destination pair and timeband.

$$\text{RBT}_{OD}^t = (TT_{95\%})_{OD}^t - (TT_{50\%}^{\text{ref}})_{OD}^t \quad (4.1)$$

- where - t and OD are respectively the timeband and O-D pair of interest
- $(TT_{50\%}^{\text{ref}})_{OD}^t$ is the reference median journey time for the O-D pair and timeband
- $(TT_{95\%})_{OD}^t$ is the 95th percentile journey time for the O-D pair and timeband.

The reference median waiting time could be either the median journey time for the day and timeband of interest or the rolling average median journey time for the same timeband over a specific number of days.

The median journey time gives a measure of the central tendency, and the RBT is an indicator of the spread of the journey time. The reason for choosing the 95th percentile and the median to calculate the buffer time is discussed in section 2.3.1. Uniman (2009) analyzed the journey time distribution for each origin-destination pair for time intervals that ranged from 15 minutes to an entire timeband. His sample size ranged from one day to an entire four-week period as defined at Transport for London, or 20 weekdays. The

calculation is done here at the O-D pair and timeband levels, for each day, in order to obtain a metric at the most disaggregate level.

The comparison of the journey time distributions for one origin-destination pair and timeband on two different days, and the definition of the reliability buffer times are shown in Figure 4-6 to illustrate the proposed methodology.

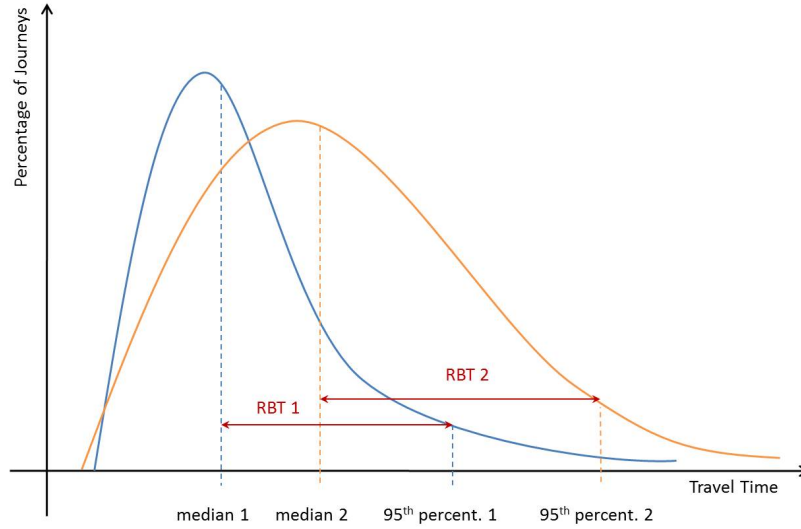


Figure 4-6: Illustrative journey time distributions for a specific O-D pair and timeband

The two curves in Figure 4-6 represent illustrative distributions of journey times on two different days, for the same O-D pair and timeband. The reliability buffer time is calculated for both days. The spread of the distribution is greater for the second day, as illustrated by the greater RBT.

The median, the 95th percentile journey time and the RBT are calculated in a similar fashion for all days at the O-D pair and timeband levels and compared across the period, or aggregated at the line level as discussed in the following section.

4.2.3 Aggregation at the Line Level

The analysis of individual origin-destination pairs is valuable for the customer since it provides information about the travel time variation for specific journeys. However, as explained before, the operator has more interest in looking at performance for each of the Underground lines rather than for individual O-D pairs independently. This section proposes a method to aggregate the O-D pair results into a line-level performance metric that

measures the evolution of journey time reliability of each Underground line independently. A metric for comparison of Underground lines will be discussed in section 4.3

The mathematical formulation of the aggregation at the line level is presented in this section. Two levels of temporal aggregation are considered here, as they may both be of interest for Transport for London:

- Underground line at the timeband level
- Underground line at the day level.

The performance of the lines during the morning and evening peaks is critical as these are the periods when the system is most heavily used. The aggregation at the day level will give a more global view of Tube journey time reliability.

The line level aggregate reliability metrics can be calculated in two possible ways. These two methods are presented and discussed in the following paragraphs.

Average median and 95th percentile journey times. The line-level median and 95th percentile are calculated as the arithmetic means of the median and 95th percentile journey times calculated for all origin-destination pairs on the line of interest. The mathematical formulation for this method is the following:

$$(TT_{\alpha})_L^T = \frac{1}{n} \frac{1}{p} \sum_{t \in T} \sum_{OD \in L} (tt_{\alpha})_{OD}^t \quad (4.2)$$

- where
- t represents the timebands included in time period T (one timeband if T is a timeband, three is T is an entire day)
 - $OD \in L$ are the origin-destination pairs on line L
 - n is the number of OD pairs on the line of interest
 - p is the number of time bands that are considered in the aggregation ($p = 1$ if the metrics is aggregated at the timeband level, $p = 3$ for a metric at the day level)
 - $(tt_{\alpha})_{OD}^t$ is the α^{th} percentile journey time for one O-D pair and timeband.

This method of aggregation does not involve any weighting of the percentiles. The lightest O-D pairs in terms of passenger flows have the same weight as the heaviest ones on the line.

Weighted average median and 95th percentile journey times. The line-level median and 95th percentile are calculated at the weighted median and 95th percentile calculated for each of the O-D pairs on the line. The weights are the passenger flows between two Underground stations on the line during each time band.

The general formula for the aggregation of the journey time reliability metrics at the line level is presented in Equation 4.3. For line L and the time period of interest T (either a timeband or an entire day), a weighted average α^{th} percentile journey time is given by:

$$(TT_{\alpha})_L^T = \frac{\sum_{t \in T} \sum_{OD \in L} (tt_{\alpha})_{OD}^t \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in L} f_{OD}^t} \quad (4.3)$$

- where
- t represents the timebands included in time period T (one timeband if T is a timeband, three if T is an entire day)
 - $OD \in L$ are the origin-destination pairs on line L
 - $(tt_{\alpha})_{OD}^t$ is the α^{th} percentile of the travel time distribution for the origin-destination pair OD and timeband t
 - f_{OD}^t is the passenger flow between origin O and destination D during timeband t .

The heaviest origin-destination pairs on the line in terms of passenger volumes are therefore weighted more. This aggregation method also gives more weight to the heaviest timebands. The AM and PM peaks see an increase in the volume of passengers who travel; however, the off-peak period is longer than the peaks and therefore the volumes are higher. The off-peak period could therefore be weighted more than the peaks for certain O-D pairs. However, this method analyzes each origin-destination pair independently and each pair is weighted according to its importance on the line.

This method of aggregation is chosen in this thesis.

The aggregate reliability buffer time for line L and time period T is simply:

$$RBT_L^T = (TT_{95\%})_L^T - (TT_{50\%}^{\text{ref}})_L^T = \frac{\sum_{t \in T} \sum_{OD \in L} \left[(tt_{95\%})_{OD}^t - (tt_{50\%}^{\text{ref}})_{OD}^t \right] \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in L} f_{OD}^t} \quad (4.4)$$

- where
- $(TT_{50\%}^{\text{ref}})_L^T$ is the weighted average reference median journey time calculated for line L with Eq. 4.3
 - $(TT_{95\%})_L^T$ is the weighted average 95th percentile journey time
 - $(tt_{50\%}^{\text{ref}})_t^{OD}$ is the reference median travel time for one O-D pair and timeband
 - $(tt_{95\%})_t^{OD}$ is the 95th percentile travel time for one O-D pair and timeband.

In a similar fashion as for the origin-destination pair and timeband analysis, the median travel time and the reliability buffer time (or the 95th percentile) are presented simultaneously to give a measure of the central tendency and the spread of the journey time distribution.

Data Selection

The analysis at the O-D pair level considered all the origin-destination pairs in the Underground system, including journeys with one (or more) transfer(s) and trips that involve a path choice.

For trips that involve an interchange, the contribution of the individual lines to the overall deterioration of the journey time reliability cannot be determined. The absence of intermediate Oyster card validation at the transfer station makes it impossible to know what leg of the trip saw an increase in its travel time. The journeys from Canary Wharf to Oxford Circus presented at the beginning of this chapter experienced an increase in RBT on Wednesday, November 16. A comparison of the RBT value and journey time distribution with those for the Canary Wharf to Bond Street trips on the same day could reasonably lead to attributing the RBT increases to the delays on the Jubilee line. However, it is not certain that the Jubilee line is the only cause of the higher RBT for the Canary Wharf to Oxford Circus journeys, since the trip involves other Underground lines.

Furthermore, a customer traveling between Canary Wharf and Oxford Circus must choose one path among the possible options described earlier. A disruption on the Victoria, Central or Bakerloo lines, if known by the passenger, would most likely affect this path choice and therefore would not necessarily affect the journey time. The journey from Canary Wharf to Oxford Circus cannot be attributed to either the Victoria, Central or Bakerloo lines, since it is not known which path the passenger chose.

Some origin-destination pairs offer two direct options, and similarly there is no way to

know what line was used by the customer. An example of this situation is the Finsbury Park to Green Park journey, shown on Figure 4-7, for which customers have two direct options: they can choose the Victoria or the Piccadilly line. The option suggested by the TfL journey planner is the Victoria line, given that the travel time on the Victoria line is generally less than on the Piccadilly line. Customers with no knowledge of the system are also likely to choose the Victoria line given that the number of intermediate stops between the two stations is smaller than on the Piccadilly line. However, according to the 2010 Rolling Origin-Destination Survey (RODS) conducted by Transport for London, 10% of the customers traveling between Finsbury Park and Green Park chose the Piccadilly line. The only information provided by the Oyster card data are the entry and exit stations and times, and it is not possible to assign this travel time to one of the lines.

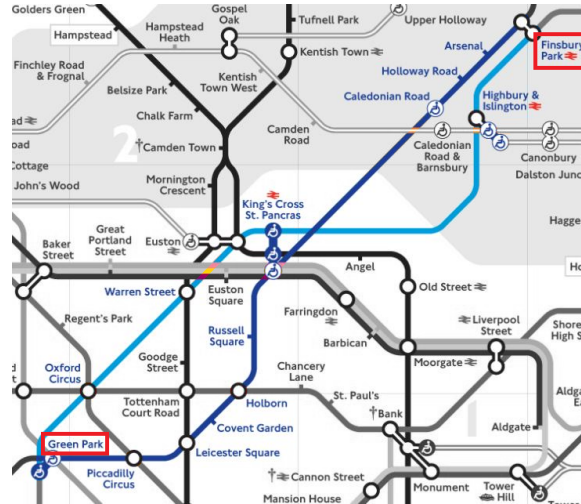


Figure 4-7: Finsbury Park to Green Park

All the journeys that involve a path choice and/or interchanges are excluded from the dataset for the aggregation at the line level. It is assumed that the customers who use a specific line as part of a multi leg journey in the Underground, will encounter the same travel conditions as the customers whose journey occurs on one line only. Passengers traveling from station A to station C via station B should encounter the same conditions between A and B as the customers whose journey starts at station A and ends at station B. The fact that the people who are transferring from one line to another are not considered for the aggregation at the line level might however bias the aggregation since the weighting is done considering the flow of passengers traveling on the line. This will be a greater issue when considering pairs which origins and/or destinations are major interchange stations

(e.g. Oxford Circus, Bank/Monument), where the number of people entering / exiting the station is comparable to the number of people transferring. However, this issue is left for further research.

4.2.4 Definition of the Journey Time Reliability Standards

So far, the proposed metrics only analyze one day at a time and show the evolution of the journey time reliability from one day to another, but do not set any standard in terms of what a good journey time should be for each origin destination pair and timeband. It is interesting and important to establish a reference in terms of what the “normal” travel time should be, and what should be expected by the customers when they travel in the Underground, according to the time of day. The journey time reliability standards should be of two types:

1. a “good” journey time distribution must be determined for each origin-destination pair and timeband in order to provide a basis for comparing all other distributions against the “good” one;
2. an “expected journey time” and an “expected buffer time” must be defined for each origin-destination pair to provide to the customers information about their journey times under “good” conditions.

This section attempts to define some standards for an assessment of the daily performance. First, the definition of a “good” day is discussed, and a methodology for the determination of a standard journey time distribution is proposed for each origin-destination pair and timeband. Second, the journey time distribution for every day is compared to the “good” one to define additional metrics for journey time reliability.

What Is a “Good” Day?

A “good” day in the Underground network can be defined in different ways, and the definition will depend on the stakeholder. For the Underground operator, a “good” day can be one with no serious incidents, or a day when all trains run on schedule. For the customers, the definition of a “good” day can be subjective, and depends on personal preferences and experiences. For this work, a “good” day is a *reliable* day and the determination of a “good” day is based on journey time reliability.

The same day is not necessarily considered as a good one for all the Underground lines. An incident can occur on one line without disrupting other lines; this would thus be a “bad” day in terms of reliability performance for the first line, but could still be a “good” day for other lines. Additionally, for the same line, the morning peak can happen without any perturbation, but the evening peak could be terrible if an incident occurs at 4 pm. Delays can also occur in one direction only. For a journey that involves transfers, a “good” day would be one without incidents on all the lines used for that trip. The “good” day should therefore be different for every line in the Underground network, for every timeband and for every origin-destination pair in the system.

The proposed approach is based on previous work by Uniman (2009). Uniman uses the 95th percentile journey time as an “indicator of delay” to distinguish between what he calls “recurrent performance” and “incident-related performance”. Each day is classified in one category or the other using a stepwise regression. However, a simpler method is adopted here in order to replace the tedious stepwise regression.

As explained above, the goal of this section is to define a good day in terms of journey time reliability. It is proposed to use the reliability buffer time to distinguish between reliable and unreliable days. The RBT simply measures the span of the journey time distribution and may not reflect incidents on the Underground if the median and the 95th percentile journey time both increase simultaneously. However, the median is not as sensitive to disruptions as is the 95th percentile journey time. As the figures will show in section 4.4, in the event of an incident, the 95th percentile journey time increases when the median generally stays constant. If the median journey time increases as well as the result of severe delays, the increase of the 95th percentile is such that the reliability buffer time will also increase. Therefore, an increase in RBT is generally indicative of disruptions.

Sample Size and Time Frame for the Determination of Standards

The sample size in terms of the number of days of data that should be considered to define a journey time reliability standard, as well as how often the standard should be updated, depend on the purpose of the metrics. Three types of journey time reliability standards may apply:

- for the publication of a daily metric, the performance for the day of interest should be compared to performance of some previous days;

- for the publication of the Oyster reliability metrics at the end of every four-week period, a “good” day and its associated journey time distribution can be defined for the period and each other day of the period is compared to the “good” one;
- for the definition of the “expected journey time” and “expected buffer time“, a larger sample size can be used.

A four-week period, or 20 weekdays, which is the reference period at Transport for London for reporting performance, is appropriate as the minimum sample size for the definition of the standards.

A seasonality effect might also be needed. The journey time might inherently vary from one period (or season) to another⁵, given the differences in weather, passenger volumes, passengers types (whether the passengers are familiar with the system could affect their travel times), etc... For instance, one could expect travel conditions to be very different in September than in August, given that the proportion of tourists unfamiliar with the network or the total number of riders will be different during these two months. Therefore, comparing performance in period 7 with that of period 6 might not be appropriate. Similarly, winter weather might have an impact on travel conditions, and comparing performance for journeys in September with performance for trips made in December might not be appropriate.

If there is a significant difference in travel conditions, the performance on one day cannot be compared to the performance in the previous period or season. The seasonality of the Oyster journey time reliability metrics will be discussed in section 4.4, and the appropriate sample size for the reliability standard will be defined in the same section. In general, standards defined based upon the largest sample size will be more robust.

Methodology for the Determination of a “Good” Day

The method proposed here is purely empirical, and tries to define a similar threshold between “good” and “bad” days that can be applicable for all origin-destination pair and all timebands. From the journey time distribution for each O-D pair, each day, and each timeband, a reliability buffer time is calculated as explained in section 4.2.2. An O-D and

⁵The seasons can be defined as followed for year 2011/2012:

- Spring: Periods 1, 2 and 3 (from April 3 to June 27, 2011)
- Summer: Periods 4, 5 and 6 (from June 26 to September 17, 2011)
- Fall: Periods 7, 8 and 9 (from September 18 to December 10, 2011)
- Winter: Periods 10, 11, 12 and 13 (from December 11, 2011 to March 31, 2012)

timeband level RBT distribution is then obtained for the appropriate time period. The graphical analysis of these RBT distributions will allow the determination of the threshold.

The threshold between “good” and “bad” days should neither be too high nor too low. A high threshold would set reliability targets that are easily reachable and therefore would not motivate the operator to improve the performance of a Tube line or the network as a whole. A low threshold on the contrary could set unrealistic targets that can be reached only on an excellent day and not under “normal” conditions. An appropriate threshold will be defined in section 4.4.4, based on journey time reliability results obtained for several O-D pairs and lines.

Once the set of good days over the time period of interest is defined for each O-D pair and timeband, a journey time distribution, referred to as *the good journey time distribution* can be picked to represent the “expected” journey conditions on a good day for this O-D pair and timeband.

Comparison of the Actual Performance With the Good Journey Time Distribution

Once a good journey time distribution has been defined for every O-D pair and every timeband, the actual journey time distributions can be compared to the good one. The comparison is proposed at two levels: at the O-D pair and timeband level and at the line and day level. The Oyster metrics previously defined are compared with the good ones, meaning that the median and 95th percentile journey times for every day are compared with the good ones. For the line level metrics, the aggregation of the good metrics from the O-D pair and timeband level to the day and line level is done in a similar fashion as the aggregation of the actual metrics, discussed in section 4.2.3. Equation 4.5 calculates the weighted average median and the weighted average of the 95th percentile of the good journey time distributions. For line L and time period of interest T (either a timeband or an entire day), a weighted average of the α^{th} percentile of the good journey time distributions is given by:

$$(TT_{\alpha})_{L,\text{good}}^T = \frac{\sum_{t \in T} \sum_{OD \in L} (tt_{\alpha})_{OD,\text{good}}^t \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in L} f_{OD}^t} \quad (4.5)$$

- where - $(tt_\alpha)_{OD, \text{good}}^t$ is the α^{th} percentile of the good journey time distribution for the origin-destination pair OD and timeband t
- f_{OD}^t is the flow of passenger between origin O and destination D during timeband t .

At the O-D pair and timeband level, the actual median journey time is compared to the median of the good journey time distribution, and the actual 95th percentile journey time to the 95th percentile of the good distribution. At the line and day level, the weighted average median and the weighted average 95th percentile of the actual journey time distributions are compared to the weighted average median and the weighted average 95th percentile of the good distributions respectively, as defined by Equation 4.5

Based on the determination of a standard travel time distribution for each origin-destination pair and timeband, a new metric is defined. For each O-D pair and timeband, the *percentage of passengers traveling in Δt or more over the good journey time* $(P_{tt \geq \Delta t})_{OD}^t$ is calculated by comparing the travel time distribution for each of the days of the period with the standard distribution. Each of the percentiles of the studied day is compared to the corresponding percentile of the standard distribution.

For each origin-destination pair, day and timeband, a set $(A_{tt \geq \Delta t})_{OD}^t$ of percentiles such that the difference between the percentile of the actual distribution and the corresponding percentile of the good distribution is greater than or equal to Δt is defined as follows:

$$(A_{tt \geq \Delta t})_{OD}^t = \left\{ \alpha \in \{1, \dots, 100\} \mid (tt_\alpha)_{OD, \text{actual}}^t - (tt_\alpha)_{OD, \text{good}}^t \geq \Delta t \right\} \quad (4.6)$$

- where - $(tt_\alpha)_{OD, \text{actual}}^t$ is the α^{th} percentile of the travel time distribution for the origin-destination pair OD and timeband t for the day of interest
- $(tt_\alpha)_{OD, \text{actual}}^t$ is the α^{th} percentile of the good travel time distribution for the origin-destination pair OD and timeband t .

The new metric $(P_{tt \geq \Delta t})_{OD}^t$ is defined as the cardinal of set $(A_{tt \geq \Delta t})_{OD}^t$:

$$(P_{tt \geq \Delta t})_{OD}^t = |(A_{tt \geq \Delta t})_{OD}^t| \quad (4.7)$$

The definition of the threshold Δt requires further discussion. This threshold can be considered as an *acceptable level of delay*. However, the *acceptable* delay is subjective, and will differ across customers and according to the length of the journey. If the expected travel time is 10 minutes, 5 additional minutes increases the travel time by 50%, whereas 5 minutes add only 20% to an expected journey time of 25 minutes. A 5 minute additional travel time for a journey from Heathrow 1,2,3 to Oxford Circus that takes about 55 minutes with the Underground could be considered as acceptable when it would not be acceptable for a journey from Brixton to Oxford Circus that takes about 15 minutes on the Victoria line. Proposing a threshold expressed as a percentage of the expected travel time might therefore be appropriate. However, using a percentage as the threshold could complicate the calculation of the metric and reduce its understandability. The final decision of the threshold Δt is left to Transport for London.

This new metric is probably the most meaningful one for Transport for London, as it translates the travel time metrics previously defined into a percentage of passengers affected by the perturbations. This percentage metric will show the extent of an incident and the number of passengers affected. The metric is relevant and understandable for both the operator and the customers.

4.3 Journey Time Reliability Measurement for Comparison Across Lines

From an operational perspective, studying the journey time reliability of the Tube lines independently is important in order to gain insight into the performance of each line, and to follow the evolution of the metrics. However, it is also relevant to compare different Tube lines, and possibly learn from the results of one line to improve performance of another.

The eleven lines of the London Underground system have very different characteristics, including lengths, average train speeds and spacing between stations. In order to properly compare Underground lines, the median journey time and the reliability buffer time might not be the adequate indicators.

The journey time reliability metrics defined so far do not take into account the length of the trips on each Underground line. The average journey time on a long line is likely

to be higher than the average journey time on a shorter line. Similarly, as illustrated by the trip from Heathrow to Hammersmith in the introduction of this chapter, the spread of the journey time distribution is usually greater for longer trips, and therefore the reliability buffer time will be greater.

It is proposed to normalize the reliability buffer time metric by dividing it by the median journey time, in order to account for the length of the average trip. The Normalized Reliability Buffer Time (NRBT) is defined for each origin-destination pair and timeband as:

$$\text{NRBT}_{OD}^t = \frac{\text{RBT}_{OD}^t}{(TT_{50\%})_{OD}^t} \quad (4.8)$$

- where
- $(TT_{50\%})_{OD}^t$ is the median journey time for the O-D pair and the timeband of interest
 - RBT_{OD}^t is the reliability buffer time calculated in Equation 4.1

From this O-D pair and timeband normalized reliability buffer time, an aggregate line-level metric is calculated for each day as:

$$\text{NRBT}_L^T = \frac{\sum_{t \in T} \sum_{OD \in L} \text{NRBT}_{OD}^t \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in L} f_{OD}^t} \quad (4.9)$$

- where
- NRBT_{OD}^t is the NRBT calculated for the origin-destination pair OD and timeband with Equation 4.8
 - f_{OD}^t is the flow of passenger between origin O and destination D during timeband t .

A small normalized reliability buffer time reflects a narrow journey time distribution, and thus a more reliable journey assuming that the median and the 95th percentile journey times do not increase proportionally.

This ratio is used for the purpose of comparing Underground lines only.

4.4 Applications

This section illustrates the use of the metrics previously defined through various applications. Section 4.4.1 describes the Oyster data processing; section 4.4.2 presents the results at the origin-destination pair and timeband level; section 4.4.3 shows the results aggregated for a single Underground line; section 4.4.4 discusses the reliability standards, section 4.4.5 illustrates the comparison of the daily performance with a rolling average, and section 4.4.6 illustrates the normalized reliability buffer time for performance comparison across lines.

4.4.1 Data Processing

The Oyster records available for this research include all LU journeys made in Periods 4 to 12 of year 2011/2012⁶, from Sunday, June 26, 2011 to Saturday, March 3, 2012.

The raw Oyster data records all the transactions made in the system, including when people top up their card or reload a monthly pass. The data of interest for this analysis consists only of the entry and exit validations for the journeys made in the Underground. It is processed in two steps. The first step links the entry and exit made with the same card in order to define a journey. This allows the calculation of the journey times between all station pairs in the system, as long as some customers traveled between them. The result of this first step is a table with each row corresponding to one journey, with the origin station, the destination, the entry and the exit times. This data can be used for a very detailed analysis, hour by hour for instance. For this thesis however, the data is aggregated to the timeband level.

The second step of the data processing determines the number of passengers who traveled between specific origins and destinations during a specific timeband and in a given journey time.

Table 4-1 gives an example of the Oyster data after it is processed. Each row shows the code of the stations of entry (852 for Canary Wharf) and exit (524 for Bond Street), the day when the journeys occurred (11640 for November 14, 2011), the journey time in minutes and the timeband. The final column gives the number of passengers who experienced that journey time when traveling between the origin and the destination on that day and timeband. In the example, 79 passengers traveled from Canary Wharf to Bond Street in 20

⁶The periods for financial year 2011/12 are defined in Appendix B.

minutes during the PM peak of Monday, November 14.

StationOfFirstEntry	StationOfExit	Daykey	JourneyTime	TimeBand	Journeys
852	524	11640	18	PM Peak	4
852	524	11640	19	PM Peak	26
852	524	11640	20	PM Peak	79
852	524	11640	21	PM Peak	97
...

Table 4-1: Example of processed Oyster records

4.4.2 Journey Time Reliability Metrics at the Origin-Destination Pair Level

This chapter was motivated with the presentation of three origin-destination pairs in the London Underground. The journey time distributions for the trips from Canary Wharf to Bond Street, from Heathrow (Terminal 1,2,3) to Hammersmith and from Canary Wharf to Oxford Circus were shown on Figure 4-3 (page 60). The Oyster journey time reliability metrics for the three O-D pairs are presented and discussed in this section.

Oyster Journey Time Reliability Metrics

Table 4-2 shows the median and the 95th percentile journey times for the three origin-destination pairs, for the PM peaks for all days of period 9. The median journey times for the Central London trips (from Canary Wharf to Bond Street and Oxford Circus) vary very little, typically within ± 1 minute. The variability of the median journey time is greater for the journeys from Heathrow to Hammersmith, as observed previously in the travel time distributions in Figure 4-3b. This can in part be explained by the longer headways on the branch of the Piccadilly line compared to the Jubilee line.

For all O-D pairs, the reliability buffer times are more variable. It is interesting to note that the incidents that occur on the Tube lines result in an increase in the reliability buffer time, while the median journey times remain similar, as shown in Table 4-2. Increases in the median journey time are generally caused by severe delays or even suspensions of the lines of interest. Maintaining a constant median journey time in the Underground is critical for Transport for London. A consistent increase in the median journey times would be a strong indicator of an overall deterioration of the service provided in the Tube. It is therefore

	Canary Wharf to Bond Street			Heathrow Terminals 1,2,3 to Hammersmith			Canary Wharf to Oxford Circus		
Day	Journeys	Median (min)	RBT (min)	Journeys	Median (min)	RBT (min)	Journeys	Median (min)	RBT (min)
14-Nov	425	22	4	116	39	7	80	25	5
15-Nov	538	22	3	99	42	13	141	26	4
16-Nov	586	23	8	95	38	5	142	27	14
17-Nov	867	23	4	125	38	4	177	26	7
18-Nov	654	22	3	116	39	8	166	26	5
21-Nov	431	23	10	113	38	5	93	28	13
22-Nov	537	23	3	88	38	5	106	26	5
23-Nov	675	23	4	94	39	13	134	29	6
24-Nov	785	23	3	110	37	5	149	26	7
25-Nov	801	24	5	117	37	6	163	27	9
28-Nov	461	22	3	106	41	12	112	26	5
29-Nov	594	23	2	99	40	6	135	25	6
30-Nov	761	23	5	93	37	6	169	26	8
01-Dec	756	23	4	114	40	6	207	27	6
02-Dec	717	23	6	124	38	8	159	27	6
05-Dec	615	23	5	109	36	4	119	26	7
06-Dec	590	22	4	109	38	6	120	26	6
07-Dec	736	23	9	90	36	5	185	27	9
08-Dec	859	23	4	103	38	4	178	26	5
09-Dec	666	23	16	124	40	5	204	28	14

Table 4-2: Journey time reliability metrics for 3 O-D pairs - PM peak - Weekdays of Period 9

important and valuable for TfL to monitor the evolution of both the median journey time and the reliability buffer time.

Seasonality in the Journey Time Reliability Metrics

Seasonality effects in the journey time distribution could be expected in the London Underground. The travel conditions in the system can be different according to the season and can be affected by multiple factors. The weather and the ridership among others can affect the journey time in one way or another. This section proposes to test the existence of seasonality effects at the O-D pair and timeband level, by looking at the trips between Canary Wharf and Bond Street on the Jubilee line and the trips between Heathrow Terminals 1,2,3 and Hammersmith on the Piccadilly line during the PM peak in periods 4 to 12 (period 6 excepted⁷), i.e. in the summer, fall and winter.

The seasonal variations of the median journey time and the reliability buffer time are considered; the goal is to evaluate if the median and the RBT vary from one period to another. Table 4-3 presents the average median journey times and average reliability buffer times for the eight four-week periods of interest, as well as the variance of the two metrics for the each of the periods, for the trips from Canary Wharf to Bond Street in the PM peak.

The average median journey times shown in Table 4-3 for the eight periods are all within two minutes, and thus within the ± 59 -second margin of error. Within one period, the variability of the median journey time is low, as shown by the values of the variance. The higher variances for periods 5 and 8 can be explained by bad performance on one day of each period.

There is a greater difference between the reliability buffer time values, and the variability of the buffer times within one period is significant for all periods except 7 and 11. The RBT variability is related to the incidents and delays that occurred on the Jubilee line during these six months. Periods 7 and 11 are better periods in terms of reliability on the Jubilee line, since the average RBT and the variance are smaller than for any other periods.

In general, the high variances of the median journey times and the reliability buffer

⁷Due to an error when the raw Oyster data was processed and imported into the database included in the analysis, no data was available at the O-D pair and timeband levels for period 6. However, the metrics at the line level were calculated from another data source before the raw Oyster data was lost and therefore is used for the analysis at the line level.

	<i>Weekdays</i>	<i>Median Journey Time</i>		<i>RBT</i>	
		<i>Average</i> (minutes)	<i>Variance</i>	<i>Average</i> (minutes)	<i>Variance</i>
Period 4	20	23.65	0.98	5.60	30.57
Period 5	20	23.00	1.89	7.05	84.68
Period 6	20	N/A	N/A	N/A	N/A
Period 7	20	22.35	0.24	4.95	3.52
Period 8	20	22.75	1.25	6.15	22.77
Period 9	20	22.80	0.27	5.25	10.83
Period 10	17	22.47	0.26	4.76	23.19
Period 11	20	22.10	0.09	4.00	2.42
Period 12	20	22.45	0.47	5.65	29.71

Table 4-3: Seasonality effects - Canary Wharf to Bond Street - PM peak - Periods 4 to 12

times are due to bad performance on a single day during the period:

- *Period 4.* The part service suspension of the Jubilee line and the delays that followed on July 5 resulted in an increase of the median journey time to 26 minutes (compared to 22 to 24 minutes for most other days of the eight months of interest) and of the RBT to 28 minutes. This incident explains the two variances for period 4.
- *Period 5.* The high variances for the median journey times and the RBT are the result of a part service suspension of the Jubilee line at the beginning of the PM peak on August 9. This incident led to severe delays throughout the entire peak period, and resulted in a 28-minute median journey time and a 44-minute RBT (by far the highest RBT of the eight periods). The high variance for the RBT is also in part explained by minor delays on several other days of the period.
- *Period 8.* The failure of the transmission-based train control (TBTC) system in the PM peak on November 4 led to a service suspension of the whole Jubilee line. This resulted in a median journey time of 27 minutes and an RBT of 24 minutes. This incident explains the variances for the median and RBT for period 8.
- *Period 10.* The variance of period 10 is the consequence of the 23-minutes RBT on December 15, consecutive to a signal failure and a sick customer that resulted in severe delays.

- *Period 12.* The high variance of the RBT on period 12 is explained by an RBT of 29 minutes on February 29 due a train failure that led to severe delays on the Jubilee line throughout the entire PM peak.

Similarly, Table 4-4 shows the average median journey times and average reliability buffer times for the same eight months, as well as the variance of the two metrics for each of the periods, for trips from Heathrow to Hammersmith.

	<i>Weekdays</i>	<i>Median Journey Time</i>		<i>RBT</i>	
		<i>Average</i> (minutes)	<i>Variance</i>	<i>Average</i> (minutes)	<i>Variance</i>
Period 4	20	38.60	8.15	8.65	18.45
Period 5	20	38.35	5.08	8.10	8.83
Period 6	20	N/A	N/A	N/A	N/A
Period 7	20	37.75	6.20	8.65	26.98
Period 8	20	37.35	2.55	5.95	2.58
Period 9	20	38.45	2.47	6.65	8.03
Period 10	17	39.29	9.10	8.59	36.51
Period 11	20	38.70	28.12	7.25	6.20
Period 12	20	39.00	8.63	9.75	49.88

Table 4-4: Seasonality effects - Heathrow Terminals 1,2,3 to Hammersmith PM peak - Periods 4 to 12

The average median journey times, as shown in Table 4-4, are within 2 minutes. The variability of the median journey times within each period is higher than for the Canary Wharf to Bond Street trip, which correlates with the observations previously made about Figure 4-3b and Table 4-2. The variance of 28 minutes for period 11 is explained by a 61-minute median journey time on January 10 because of a track failure in the PM peak between Hammersmith and Earl’s Court that resulted in severe delays on the west end of the Piccadilly line. In a similar fashion to the Canary Wharf to Bond Street trip on the Jubilee line, the variability of the RBT within each period can be explained by the incidents and delays that occurred on the Piccadilly line. However, if the high variances for the first O-D pair can be explained by one or two “bad” days with incidents that resulted in severe delays, the high variances for the second O-D pair are explained by smaller, but more numerous delays across the periods.

Tables 4-3 and 4-4 show no evidence of any seasonality effect for these two origin-destination pairs. There are differences in the average median travel times and average RBT between the eight periods that are considered; however these differences are due to the number and the severity of the incidents that occurred during each of the periods, and the incidents are not related to the season or the period in which they occurred (with the exception of the delays on the network due to the snow storms on February 5 and February 10, 2012, which had impacts that were comparable to other major incidents).

This lack of evidence to support any seasonality effect should be confirmed with data from the same periods in the following years.

4.4.3 Journey Time Reliability Metrics at the Line Level

The Jubilee line is used as an example for the calculation of the travel time reliability performance metrics at the line level. Figure 4-8 shows the aggregated performance of the Jubilee line for all weekdays in periods 7, 8 and 9.

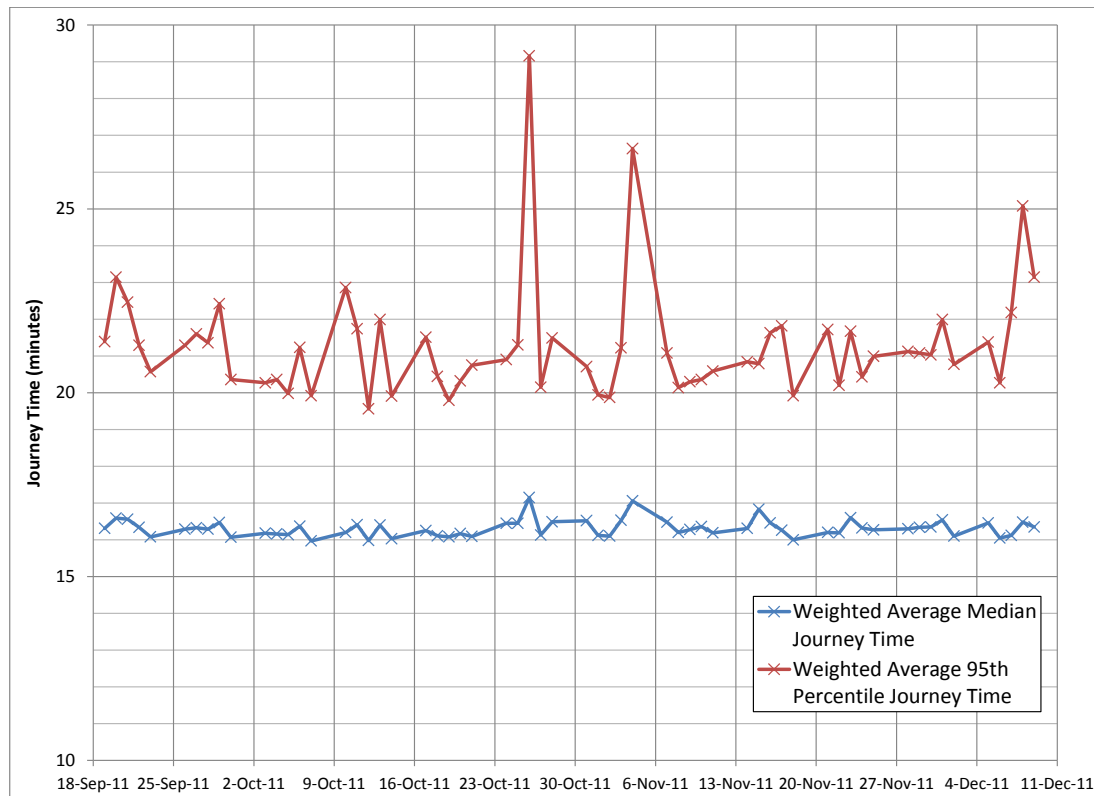


Figure 4-8: Journey time reliability metrics - Jubilee line - Periods 7, 8 and 9 - Weekdays

The graphs show a 1-minute variation of the median journey time. However, the variation of the 95th percentile and thus of the reliability buffer time, is much wider. There is a 10-minute difference between the minimum and the maximum 95th percentile journey times. The reliability buffer time reflects the disruptions on the line. For instance, the spike on October 26 is caused by a suspension of the entire Jubilee line around 19:00 and the severe delays that persisted for the rest of the evening. Similarly, the RBT peak on November 4 is due to a loss of signaling control at the end of the PM peak that triggered the suspension of the line followed by severe delays.

Test of the Seasonality Effects

The seasonality effect for the aggregate journey time reliability metrics at the line level is tested here. Table 4-5 shows the averages and variances for the aggregate Oyster metrics for the Jubilee line, for periods 4 to 12.

	<i>Weekdays</i>	<i>Weighted Average Median Journey Time</i>		<i>Weighted Average RBT</i>	
		<i>Average (minutes)</i>	<i>Variance</i>	<i>Average (minutes)</i>	<i>Variance</i>
Period 4	20	17.36	0.22	6.53	5.39
Period 5	20	16.54	0.29	6.12	4.30
Period 6	20	15.77	0.05	6.56	8.27
Period 7	20	16.26	0.03	4.93	0.81
Period 8	20	16.36	0.09	4.97	4.30
Period 9	20	16.33	0.04	5.08	1.22
Period 10	17	16.19	0.12	4.62	0.86
Period 11	20	16.16	0.06	4.85	1.15
Period 12	20	16.55	0.50	6.03	5.30

Table 4-5: Seasonality effects - Jubilee line - Periods 4 to 12

Similarly to the results at the O-D pair and timeband level, the weighted average median journey times are all within 2 minutes. The variability of the median within any period is small. The difference in RBT from one period to another is more significant, as well as the variability of the aggregated RBT within any period. However, the values do not suggest any seasonal variation of the median journey times and the reliability buffer times.

This absence of evidence supporting seasonality effects at the origin-destination pair and

timeband level as well as the line and day level, allows the journey time reliability standards to be defined for the time period that is most appropriate and provides more robust results. The performance for trips made in December can be compared to that of trips made in July, and a reliability standard could be defined for each O-D pair and timeband, using data from an entire year, and all days of the next year can be compared with an annual standard.

4.4.4 Journey Time Reliability Standards

This section presents the definition for the journey time reliability standards as discussed in section 4.2.4, based on the results of the previous section.

Reliability Buffer Time Distribution

An empirical method based on the reliability buffer time distribution was proposed in section 4.2.4 to define the threshold between “good” and “bad” days. Figure 4-9 presents the reliability buffer time distribution for each line for the Fall 2011, i.e. for periods 7, 8 and 9. The distributions are presented at the day and line level as the curves are smoother than the distributions at the O-D pair and timeband level. The RBTs at the O-D pair and timeband level are whole numbers given the format of the Oyster data used for this analysis, and therefore the RBT distributions are step functions. Aggregated at the line and day levels, weighted average RBTs are real numbers. The day and line level RBT distributions make it easier to observe slope changes. Additionally, the proposed threshold should also allow the determination of “good” and “bad” days at the line level.

The slope of the distribution changes when reliability starts to deteriorate. The first part of the graph, where the slope of the curve is steeper, corresponds to the “good” or “recurrent” days in terms of reliability. The threshold between “good” and “bad” days should correspond to the percentile at which the slope of the distribution changes.

The major slope changes for all the lines are indicated by the ovals on Figure 4-9. For all the lines except the Northern line, the first change of slope of the RBT distributions occurs between the 40th and 80th percentiles. For the Northern line, the slope changes after the 80th percentile indicating that the line is more reliable than the others. Considering the 50th percentile of the RBT distribution as the threshold between “good” and “bad” days seems reasonable and intuitive, and this value can be used for all the lines. This leads to

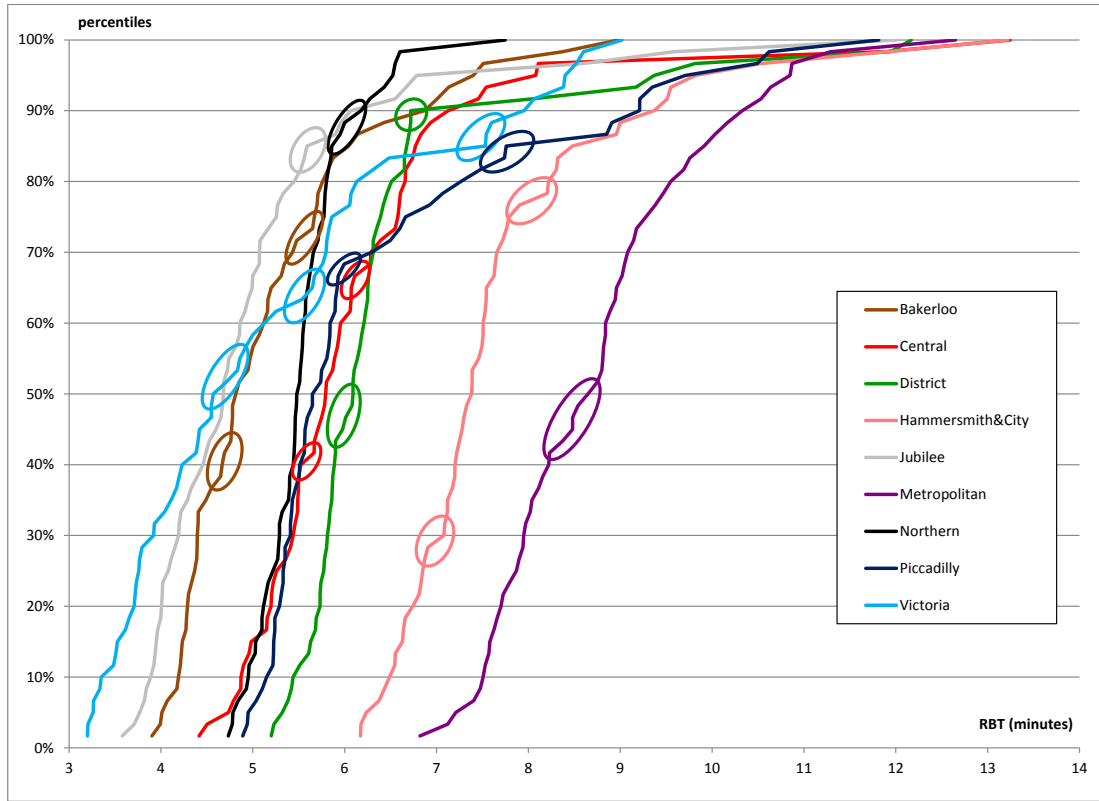


Figure 4-9: Reliability Buffer Time distributions - Periods 7, 8 and 9 - Weekdays

realistic expectations in terms of reliability standards; however, this threshold is subjective and can be changed by the agency. Increasing the threshold will lead to looser targets, while decreasing this threshold would set stronger standards, less easily achievable.

Once the threshold is set, all the days with a reliability buffer time smaller than the 50th percentile of the RBT distribution are considered as “good” days. The 25th percentile of the RBT distribution is the median RBT for the set of “good” days. The date associated with this 25th percentile can be used as the good day for comparison purposes; the journey time distribution on that day is *the good journey time distribution*.

The results in the previous section showed no evidence of seasonality effects as far as the median journey times and the reliability buffer times are concerned. The following time frames are therefore proposed for the definition of the reliability standards. For the calculation and publication of daily metrics, the performance on one day can be compared to the good day of the previous period. For the publication of the reliability metrics at the end of each four-week period, alongside the existing KPIs, a good journey time distribution can be calculated for each O-D pair and timeband using data for the entire period. For

publication to the customers of an expected journey time and an expected buffer time, data for an entire year can be used, and the numbers can be updated when the schedule changes on one of the Underground lines.

Actual vs. “Good” Journey Time Reliability Metrics

This section presents the results of the metrics defined in section 4.2.4, using the day corresponding to the 25th percentile of the reliability buffer time distribution as the reference day for comparison (good day).

Section 4.2.4 introduced a new metric, the percentage of passengers traveling in Δt or more over the good journey time $(P_{tt \geq \Delta t})_{OD}^t$. The threshold Δt is fixed to 5 minutes for this analysis. The metric is therefore called the *percentage of passengers traveling in 5 minutes or more over the good journey time*. $(P_{tt \geq 5\text{min}})_{OD}^t$ is calculated as followed:

$$\begin{aligned} (P_{tt \geq 5\text{min}})_{OD}^t = & \sum_{i=5}^{10} (P_{tt=i})_{OD}^t + (P_{11\text{min} \leq tt \leq 15\text{min}})_{OD}^t + (P_{16\text{min} \leq tt \leq 20\text{min}})_{OD}^t \\ & + (P_{21\text{min} \leq tt \leq 25\text{min}})_{OD}^t + (P_{26\text{min} \leq tt \leq 30\text{min}})_{OD}^t + (P_{tt \geq 30\text{min}})_{OD}^t \quad (4.10) \end{aligned}$$

- where
- $(P_{tt=i})_{OD}^t$ is the percentage of passengers traveling on the O-D pair and timeband of interest in exactly i minutes over the good journey time
 - $(P_{i \leq tt \leq j})_{OD}^t$ is the percentage of passengers whose travel time was between i and j minutes above the good journey time.

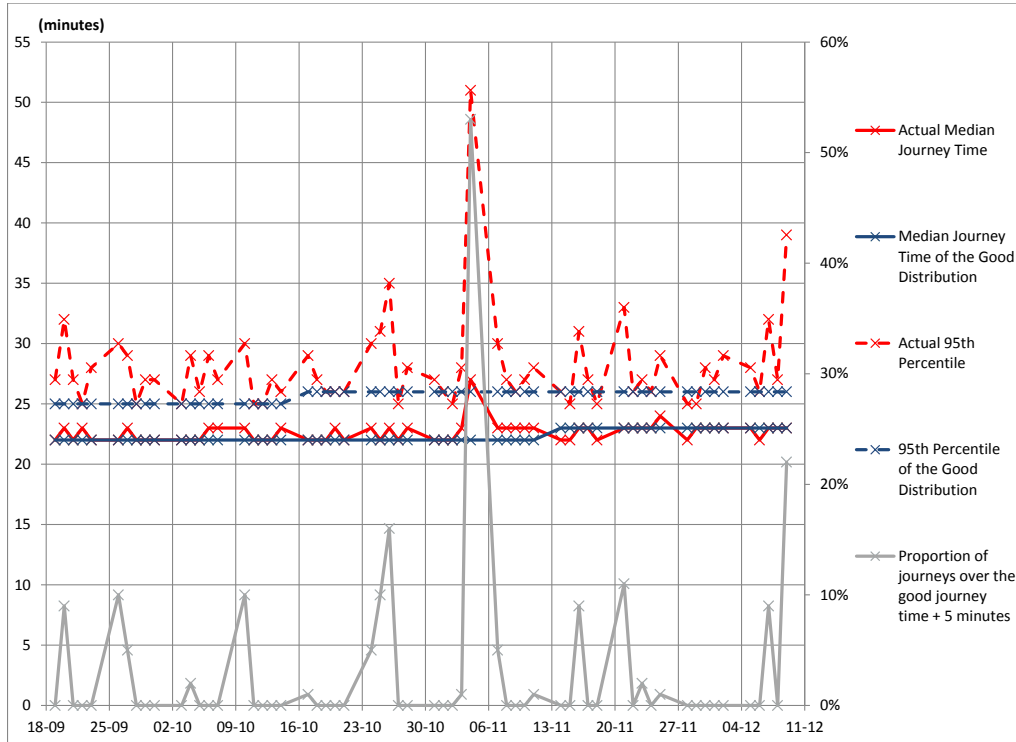
Therefore, a different threshold can be chosen and an additional metric could also be defined. The percentage of people whose travel times were more than 20 or 30 minutes over the good journey time would, for instance, reflect the significant incidents.

Origin-destination pair and timeband levels. Figure 4-10 compares the actual daily journey time reliability metrics with the metrics for the good day, for the trips from (a) Canary Wharf to Bond Street, (b) Heathrow to Hammersmith and (c) Canary Wharf to Oxford Circus, for PM peak of all weekdays in periods 7, 8 and 9. For each O-D pair, “good” median and 95th percentile journey times are defined for each four-week period, which explains why the good metrics are not constant across the three periods considered. Figure 4-10 shows a higher variability of the median and 95th percentile journey times

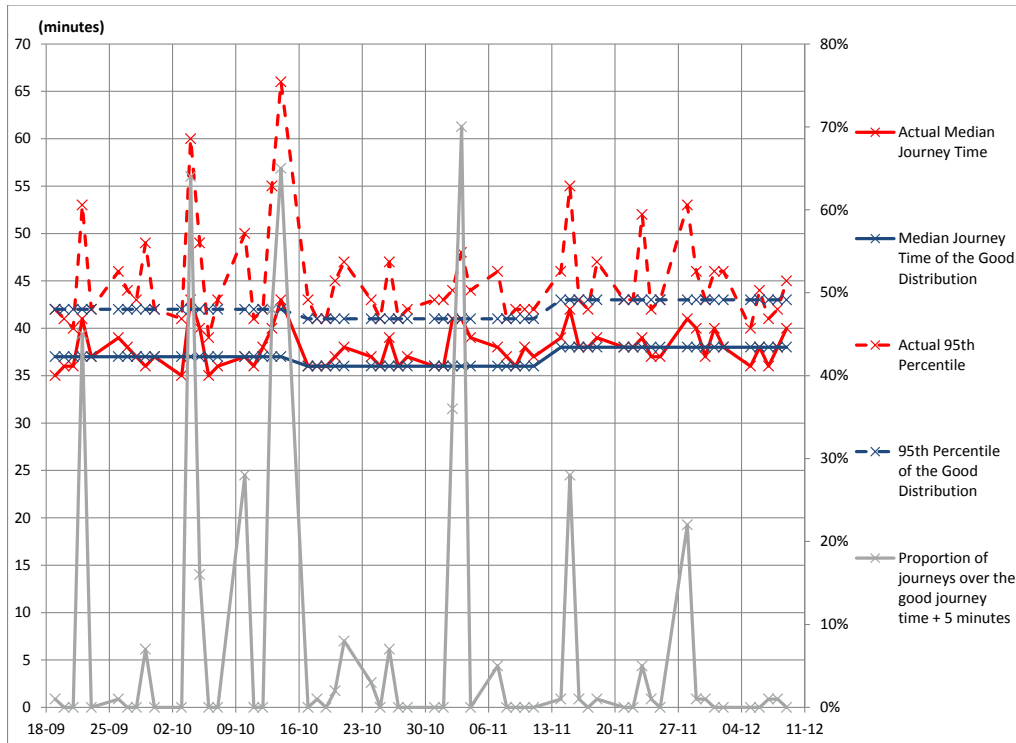
for the trips from Heathrow to Hammersmith compared to the two other O-D pairs, as previously noted in Table 4-2. The percentage of passengers traveling in 5 minutes or more over the good journey time is also more variable for the Piccadilly line trip. In general, the proportion of journeys that took 5 minutes (or more) over the good journey time is strongly correlated with the 95th percentile of the journey time distribution.

Line-level. Figure 4-11 presents the Oyster journey time reliability metrics for the Jubilee line, for weekdays in periods 7 to 9. The five metrics are aggregated at the line and day level according to the methodology defined in the previous sections. The weighted average median journey times are more stable than the median journey times at the O-D pair and timeband levels. As previously noted, there is strong correlation between the percentage of journeys 5 minutes or more over the good journey time and the weighted average 95th percentile. It is interesting to note that for all days, at least some of the journeys take longer than they should, but the proportion of journeys 5 minutes or more over the good journey time remains below 4% for most days.

The weighting method using the passenger volumes explains a simultaneous increase of the metrics on certain days. For instance, on November 15 an international soccer match took place in the evening at the Wembley Park, generating a significant increase in the passenger flow on the Jubilee line on that day. This translated into an increase of the weighted average median and 95th percentile of the good journey time distributions, as observed on Figure 4-11.

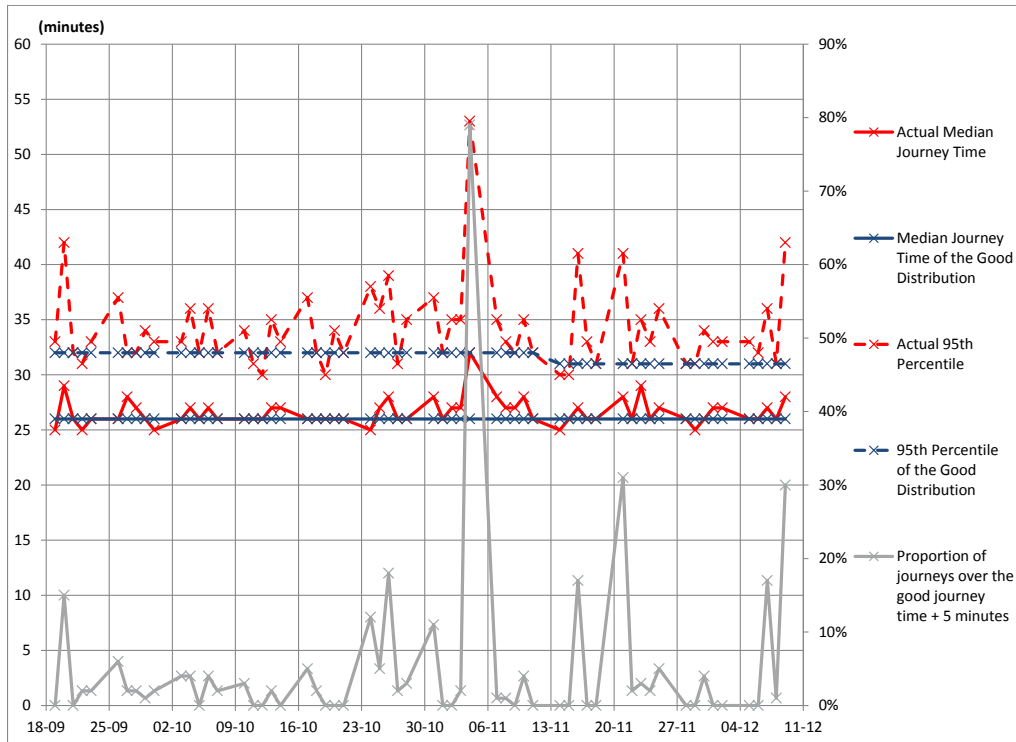


(a) Canary Wharf to Bond Street - Jubilee line



(b) Heathrow Terminals 1,2,3 to Hammersmith - Piccadilly line

Figure 4-10: Comparison of actual and good journey time reliability metrics - Three O-D pairs - Weekdays PM Peak - Periods 7, 8 and 9



(c) Canary Wharf to Oxford Circus

Figure 4-10: Comparison of actual and good journey time reliability metrics - Three O-D pairs - Weekdays PM Peak - Periods 7, 8 and 9 (Continued)

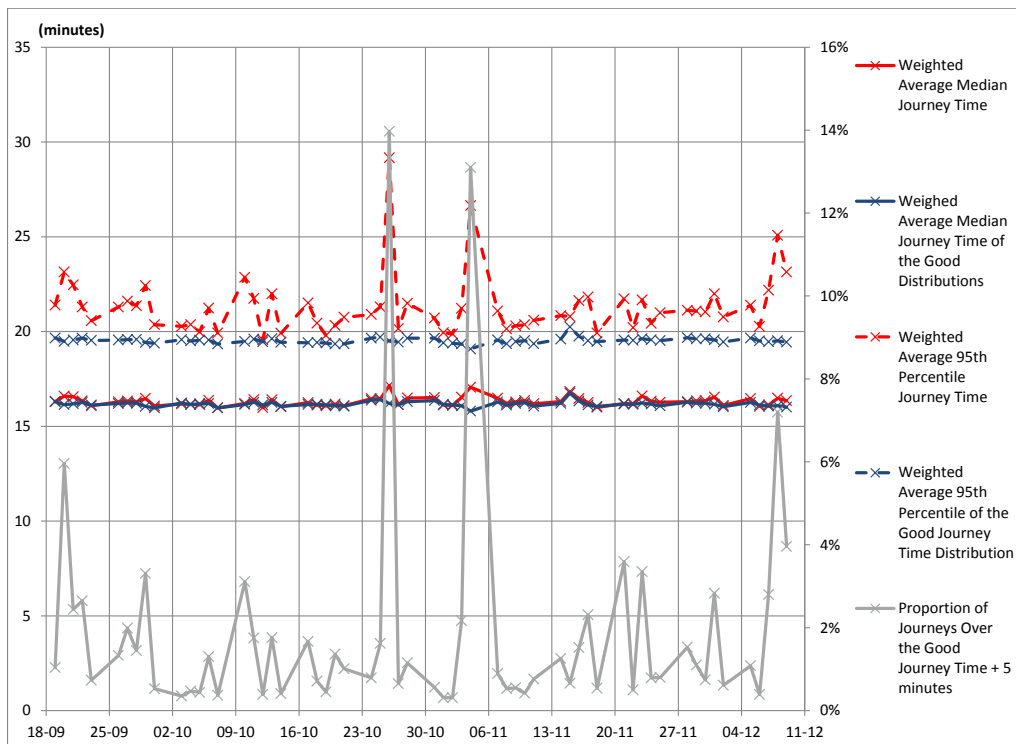


Figure 4-11: Comparison of actual and good journey time reliability metrics - Jubilee Line
- Weekdays - Periods 7, 8 and 9

4.4.5 Comparison of the Daily Performance with a Rolling Average Median Journey Time

An additional comparison is proposed in this section to illustrate the possibility of choosing a different reference median journey time. At the origin-destination pair and timeband levels, the daily performance is compared with a rolling average median journey time; at the line and day levels, the performance is compared with a rolling average of the weighted average median journey times. The rolling average is calculated over the 20 previous weekdays. For instance, the performance on October 17, 2011 (the first week day of period 8) is compared to the average calculated all the weekdays of period 7. The rolling average $(TT_{50\%})^D$ for day D is calculated using the passenger volumes as weights, as follows:

$$(\overline{TT}_{50\%})^D = \frac{\sum_{d=D-1}^{D-20} (tt_{50\%})^d \cdot f^d}{\sum_{d=D-1}^{D-20} f^d} \quad (4.11)$$

- where
- $(tt_{50\%})^d$ is the median travel time for the day of interest (either at the O-D pair / timeband level or aggregated at the line and day level)
 - f^d is the passenger flow on the O-D pair/timeband or the line/day of interest.

Figures 4-12a and 4-12b show the comparisons for the Canary Wharf to Bond Street journeys during the PM peak and for the Jubilee line respectively, for periods 8 and 9.

The rolling average allows the comparison of the daily performance with a quasi constant number. The difference between the actual 95th percentile and the rolling average characterizes the performance on a single day. This “buffer time” can be used to evaluate the extra time required each day compared on the median time required to travel on the 20 previous days, while the reliability buffer time defined previously compares the extra time required on each day to the median journey time on the same day. This comparison with the rolling average makes it possible to abstract the increases of the daily median by comparing with a more stable number.

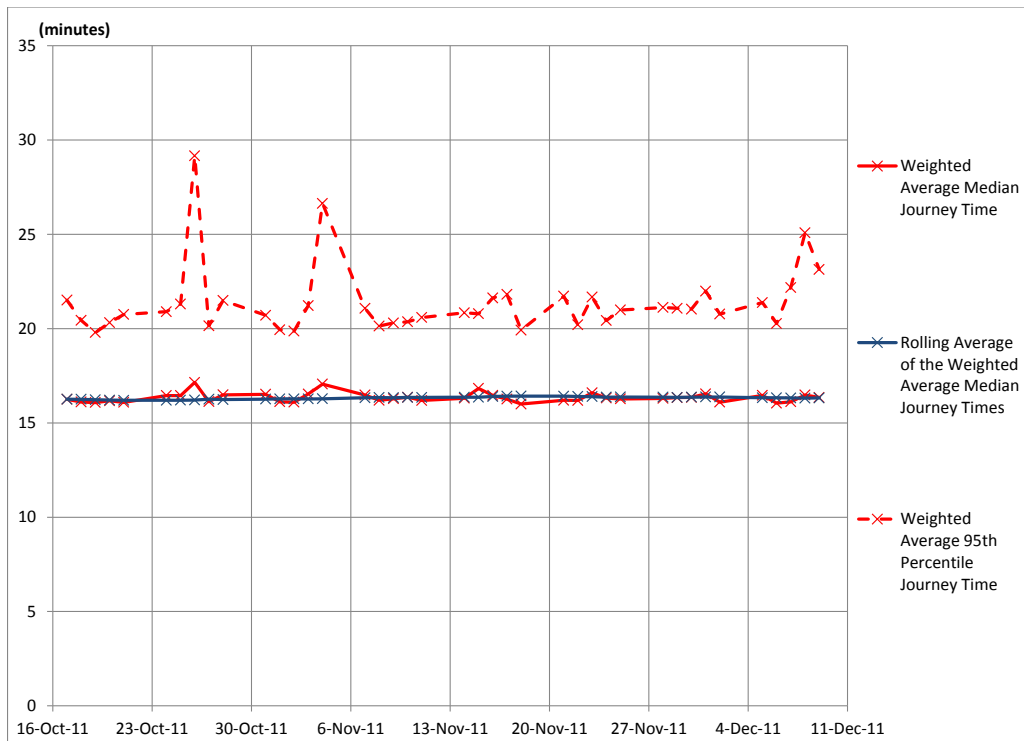
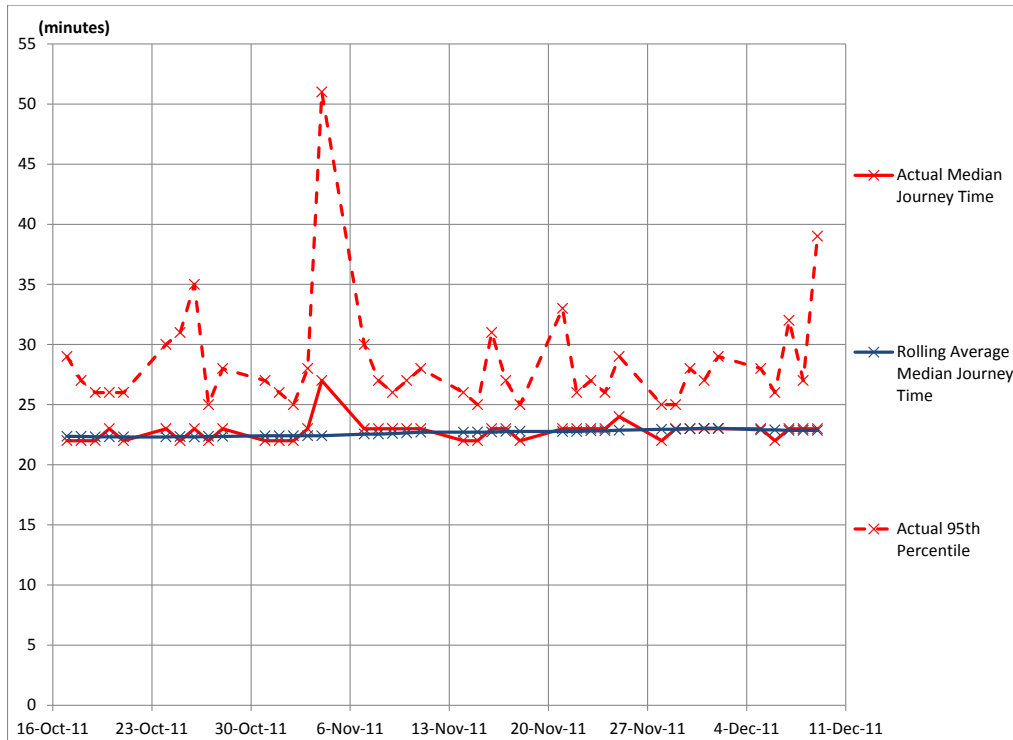


Figure 4-12: Comparison of the daily performance with a rolling average median journey time over the previous 20 weekdays

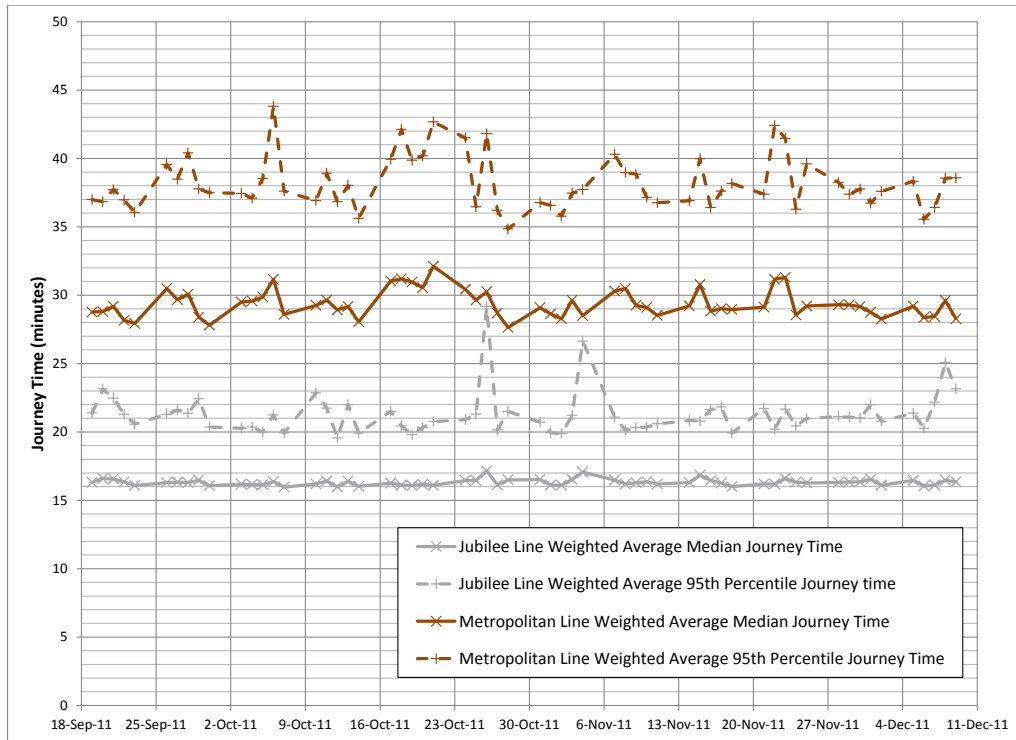
4.4.6 Metric for Comparison of Underground Lines

The comparison of journey time reliability performance of two Underground lines was described in section 4.3. The Jubilee and the Metropolitan lines are used to illustrate the comparison of performance across lines. Figure 4-13a compares the median and 95th percentile journey times for all the weekdays in periods 7, 8 and 9, for the Jubilee and the Metropolitan lines.

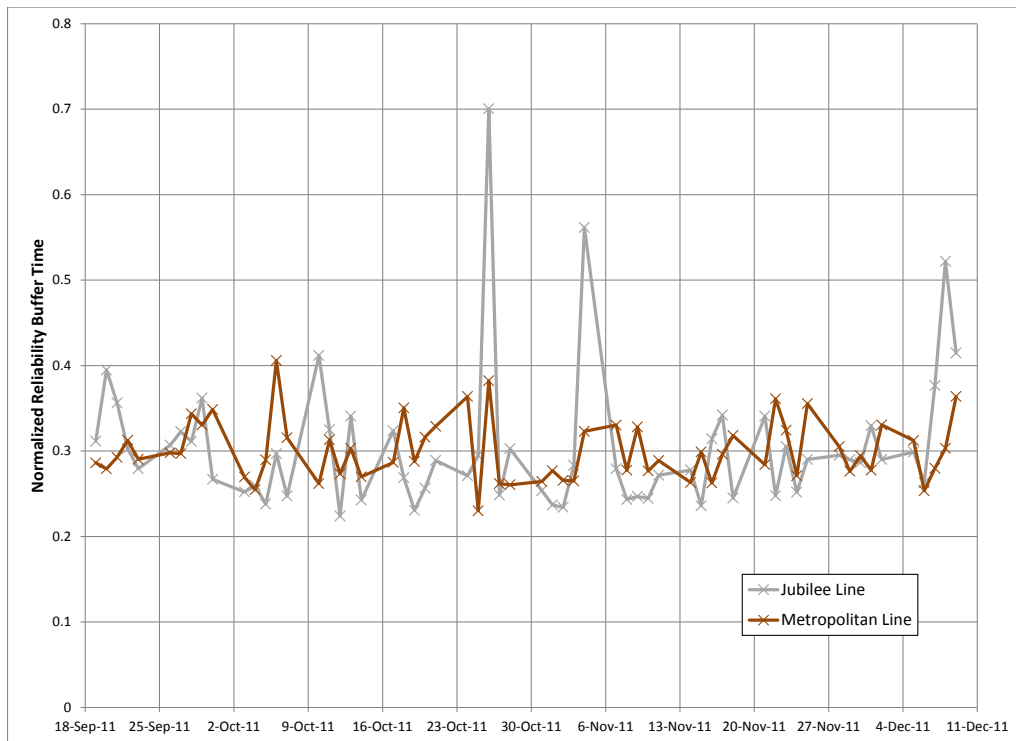
A first glance at the graph would suggest that the Metropolitan line is less reliable than the Jubilee line, as the median journey times are, on average, larger for the Metropolitan line, and the average reliability buffer time for the Metropolitan line is also larger. However, the Metropolitan line is significantly longer than the Jubilee line (41.4 versus 22.5 miles) and extends from Central London all the way to fare zone 9 in the northern suburbs. The Jubilee line serves zones 3 to through Central London. The trips made on the Metropolitan are thus, on average, longer than those made on the Jubilee line, as shown by the median journey times.

Normalizing the reliability buffer time by the median journey time allows a fairer comparison of reliability performance of these two lines, as illustrated by Figure 4-13. The average NRBT for both lines is around 0.3, meaning that on average the RBT is 30% of the median travel time. The ratio still reflects the incidents that occurred on the Jubilee line on October 26, November 4 and December 8. The magnitude of the delays on the Metropolitan line is smaller than on the Jubilee.

The normalized reliability buffer time is a good metric for comparison. However, the ratio does not give information about the variability of the median and the 95th percentile journey times across the periods. Therefore, the ratio should be used for comparison purposes but should not replace the median and 95th percentile journey time for the measure of travel time reliability for a single line.



(a) Median and 95th percentile journey times



(b) Normalized Reliability Buffer Time

Figure 4-13: Journey time reliability metrics: Jubilee line vs. Metropolitan line
Weekdays - Periods 7, 8 and 9

4.5 Use of the Reliability Metrics

As explained before, the customers want to see performance and / or predictions for their personal journey(s), whereas the operator LU and Transport for London are more interested in aggregate metrics that show performance at the line level.

4.5.1 Presentation for Internal Use at Transport for London

It is proposed to publish the new journey time reliability metrics alongside the existing Key Performance Indicators, to assess the performance of the London Underground line using the extensive source of data offered by Oyster. The existing KPIs are published by Transport for London on their website at the end of each four-week period. These reports show general results for the entire Underground network as well as line level metrics. Additionally, the internal reports also split the metrics into the different components of the journey to produce a detailed assessment.

The existing KPIs should be used for the breaking of the journey experience into the different components, and complemented with the new metrics to provide the granularity of measure (by O-D pair and timeband) that the existing KPIs do not provide. The new metrics can also be used for comparison of the performance within the four-week period, when the KPIs allow only the comparison of the periods themselves.

Although the KPIs are published at the end of every four-week period and averaged over 28 days, London Underground calculates at the end of every service day and publishes internally a set of measures in order to provide a daily assessment of the Tube performance. Two of these metrics are of interest as they directly relate to journey time reliability. The Lost Customer Hours (LCH) metric described in Chapter 3 is calculated every day from the incidents that occurred on each of the lines. Additionally, a proxy measure for the excess Platform Wait Time (PWT) is estimated. London Underground's objective is to provide the Journey Time Metric (JTM) in real time, including all journey components as described in Chapter 3. However, the signaling system in use on some of the lines does not allow automatic real time calculation of headways and running times. Moreover, the complexity of the calculation of the JTM makes it hard to calculate in real time. A proxy measure is therefore used for its easy calculation. The excess PWT is calculated from

successive train headways collected automatically at some selected stations on each Tube line, compared with the scheduled headways. For the Jubilee Line for instance, the proxy measure is calculated at Finchley Road (southbound) and Canary Wharf (westbound).

The rest of this section compares the current metrics with the proposed ones for the Jubilee lines, using data for periods 7 to 9. Figures 4-14 and 4-15 compare respectively the reliability buffer time with the proxy measure for excess PWT, and the percentage of journeys more than five minutes over the good journey time with the Lost Customer Hours. The two figures show a strong correlation between the proposed new metrics and the existing ones. Incidents and delays are reflected by increases in all four metrics.

The correlation between the new and the existing metrics is analyzed more closely in Figure 4-16 and Table 4-6. Figure 4-16 plots the proposed new metrics against the existing metrics. The strong correlation between the new and existing metrics is clear from the plots, and confirmed by the high correlation coefficient between the proxy excess PWT and the two new metrics in Table 4-6. The correlation with the LCH metric is slightly smaller, but still high.

	proxy	lch
rbt	0.8663	0.6471
percent	0.8256	0.6470

Table 4-6: Correlation coefficients between the existing and the proposed metrics
Jubilee line - Periods 7, 8, 9

There is a strong correlation between the existing metrics and the proposed ones. However, the perspective from which the new metrics is calculated and the components of the journey experience that are considered make them valuable to complement the current metrics.

- The Oyster-based reliability metrics are calculated using data produced by the customers, while the proxy measure for the excess PWT and the LCH are calculated from the operator's perspective. The percentage of journeys taking 5 minutes or more over the good journey time provide a measure of customers impacted by the delays on the Underground.
- The proxy measure for the excess platform wait time does not consider the entire journey experience, but focuses only on the waiting time at the station. Moreover,

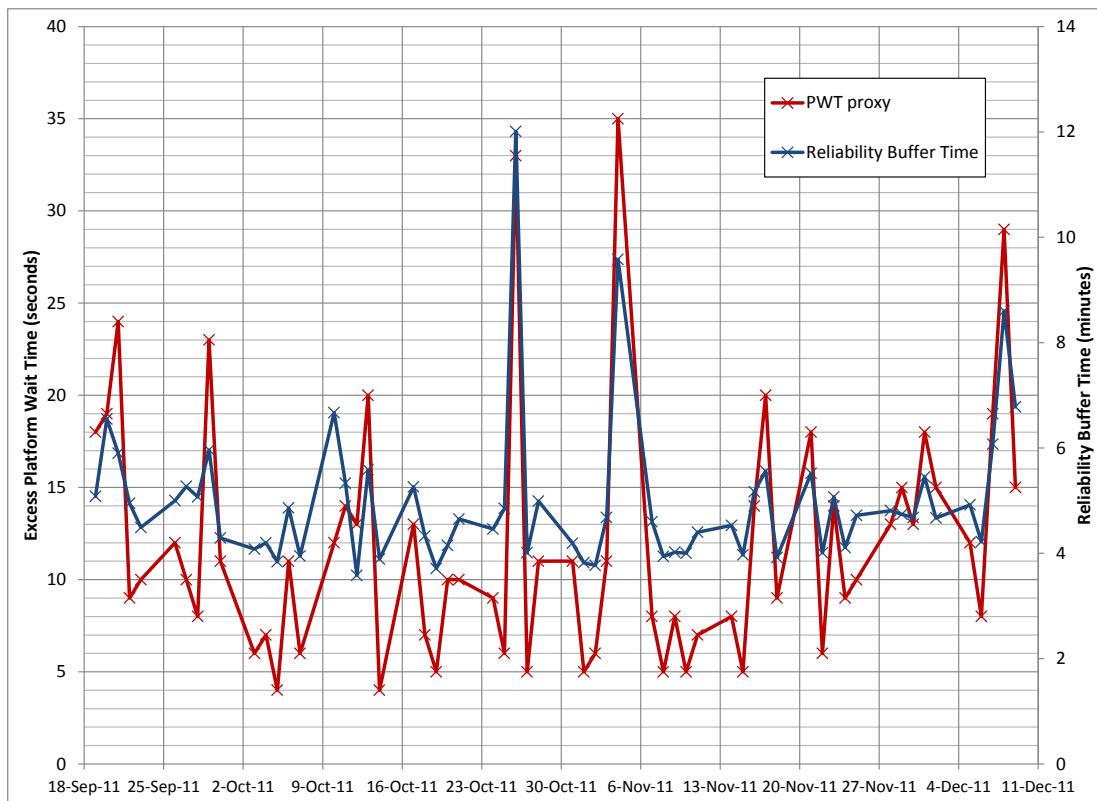


Figure 4-14: Comparison of the Reliability Buffer Time with the Platform Wait Time Proxy - Jubilee Line - Weekdays - Periods 7, 8 and 9

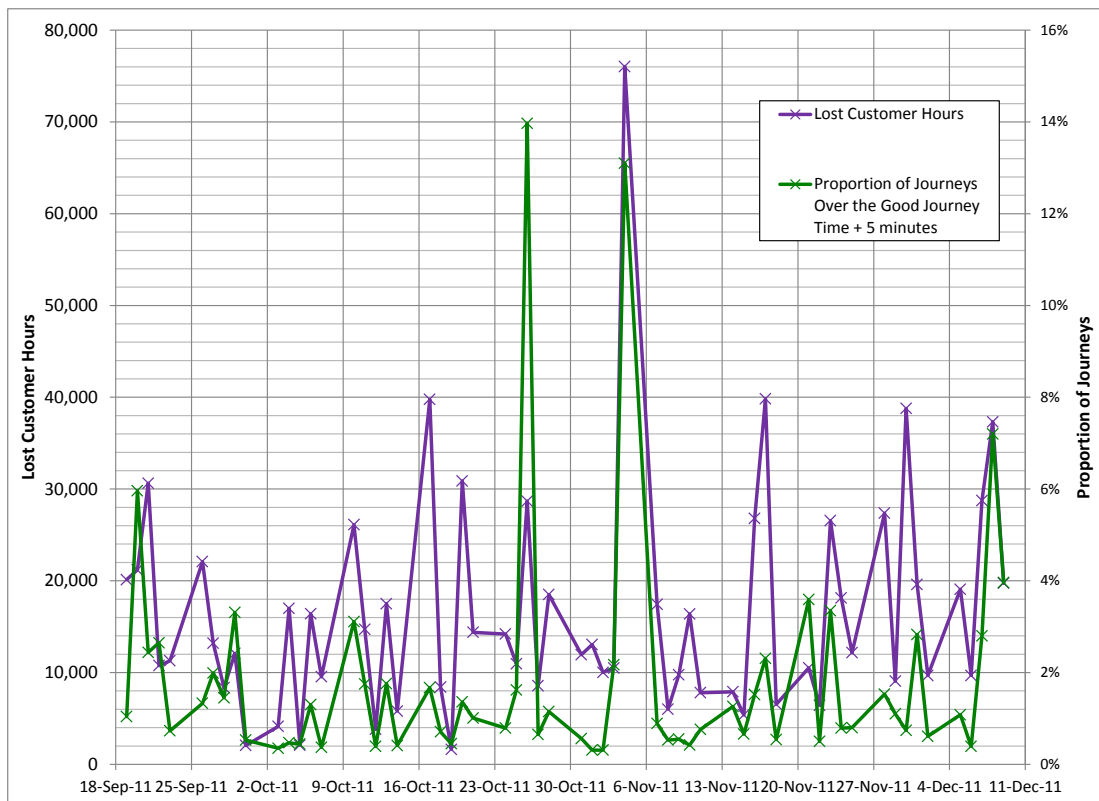


Figure 4-15: Comparison of the Percentage of journeys over the good journey time + 5 minutes with the Lost Customer Hours - Jubilee Line - Weekdays - Periods 7, 8 and 9

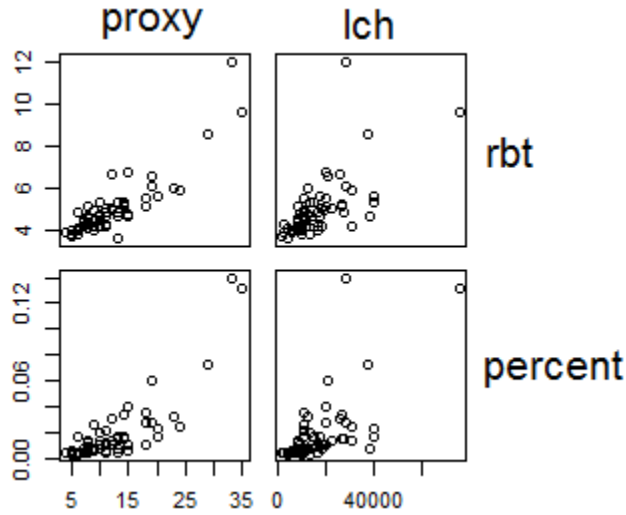


Figure 4-16: Correlation between the existing and the proposed metrics
Jubilee line - Periods 7, 8, 9

the proxy is calculated from headways recorded only at a few stations along the line, in order to simplify the calculation and allow publication daily. The RBT and the percentage of journeys taking 5 minutes (or more) over the good journey time consider the entire trip, and the calculation is based on all the Oyster records for trips made on the line of interest.

- Although it evaluates the impact of incidents and delays on customers in terms of time lost, the LCH is not a reliability metric per se, but rather a measure of the availability of service. The focus of this LCH metric is therefore different from all other indicators published by London Underground or proposed in this thesis.

The proposed Oyster journey time reliability metrics should therefore be used to complement the existing metrics and used by Transport for London and London Underground to identify actions to improve the performance of the network⁸.

⁸The proportion of journeys taking 5 minutes (or more) over the good journey time is already calculated for each line on a daily basis, using Oyster data, and is published alongside the proxy measure for excess PWT and the LCH on the Daily Dashboard on the London Underground intranet. A screen shot of this dashboard is presented in Appendix C.

4.5.2 Presentation to the Customers: Addition to the Journey Planner

As discussed before, the main goal for this analysis is to use Oyster data produced by the customers in order to provide them with understandable and meaningful metrics. The journey time reliability measures calculated in this chapter should be published alongside the existing Key Performance Indicators, and the customers should have access to these metrics of past performance. But customers also have interest in their future journey times in order to plan their trips. The journey planner is the main tool used by the Transport for London customers to plan their trips, and therefore this tool should include the reliability metrics.

Uniman (2009) proposes a possible modification of the journey planner to show the variability of journey time based on Oyster data when the journey planner currently shows a fixed journey time based on the schedule. The journey planner could give an idea of the additional time that should be budgeted by the customer in order to maximize his probability of arriving on time to his destination. On the journey planner, the formulation of this buffer time should be thought through carefully, so that it is understandable by the customers.

When planning their trips with the journey planner on the TfL website, customers are given two options: they can either request that they depart at a given time, or arrive by a given time. The proposed modifications of the journey planner for the two options, adapted from Uniman (2009), are shown on Figure 4-17.



Journey Summary							
Departing: Monday 14 November 2011							
From: Canary Wharf							
To: Bond Street							
Restrictions:							
Route	Depart	Expected Arrival	Latest Arrival	Duration (up to)	Interchanges		
1	17:00	17:22	17:26	00:22 (00:26)		View	<input checked="" type="checkbox"/>
2	16:54	17:16	17:20	00:22 (00:26)		View	<input checked="" type="checkbox"/>
earliest earlier later latest						View selected	

Figure 4-17: Proposed modification of the journey planner - Direct trip - Canary Wharf to Bond Street (adapted from Uniman, 2009)

Route 1 is the proposed change for customers who select the “Depart at” option of the journey planner, and are therefore not concerned about on-time arrival. The information displayed indicate that the customers traveling between Canary Wharf and Bond Street should expect a journey time of 22 minutes, and up to 26 minutes. The expected arrival and the latest arrival times are based on the median and 95th percentile journey time of the good distribution respectively.

Route 2 is the proposed result for customers who are concerned about on-time arrival and select the “Arrive at” option. The software would calculate the optimal departure time based on the 95th percentile journey time, and display a time that guarantees an arrival at the desired time (here, 17:20) with 95% certainty under normal travel conditions. The customers can expect to arrive by 17:16 half of the time.




However, showing journey time information based on Oyster data can be problematic for journeys that involve an interchange. The Oyster journey time is measured from entry gate to exit gate, and it does not allow the decomposition of the journey into its different stages. The journey planner currently provides information about the interchange times, based on models and estimates, and uses schedules to measure the travel time for the different legs of the journey. Figure 4-18 shows the proposed modification of the journey planner for the trip from Canary Wharf to Oxford Circus, via Bond Street.

As proposed for a direct trip, information about the variability of travel time is displayed, using the median and the 95th percentile journey times as references to provide “expected” and “maximum” journey times. The proposed modification includes an average interchange time similar to the one currently displayed by the journey planner. However, the proposed presentation does not provide travel time information for each leg of the journey, which might be criticized by the customers. This presentation can be complemented by scheduled on-train times, as calculated currently by London Underground, but this would only give an average time for each leg.

Displaying the reliability buffer time on the journey planner would be a sign from Transport for London that they acknowledge the fact that the same trip will not take exactly the same time every day, even under good conditions. It would also reduce the inconvenience for passengers.

Journey Summary

Departing: Monday 14 November 2011 at 17:00
From: Canary Wharf
To: Oxford Circus
Restrictions:

Time	Details	Information
Depart: 17:00 	<div> <div>start</div> <div>Canary Wharf Underground Station</div> <div> Take the Jubilee Line towards Stanmore Underground Station or Jubilee Line towards Willesden Green Underground Station or Jubilee Line towards Wembley Park Underground Station </div> <div>Zone(s): 2, 1</div> </div>	<div> Average interchange time: 4 mins  </div>
	<div> <div>Bond Street Underground Station</div> <div> Take the Central Line towards Hainault Underground Station or Central Line towards Debden Underground Station or Central Line towards Newbury Park Underground Station or Central Line towards Epping Underground Station or Central Line towards Loughton Underground Station </div> <div>Zone(s): 1</div> </div>	
Arrival Expected: 17:26 Latest: 17:32	<div> <div>end</div> <div>Oxford Circus Underground Station</div> </div>	

Expected journey time: 00:26
Maximum journey time: 00:32
Interchange: 1

Time	Details	Information
------	---------	-------------

Depart:

17:00

start

Canary Wharf Underground Station

Take the Jubilee Line towards
Stanmore Underground Station
or Jubilee Line towards Willesden
Green Underground Station
or Jubilee Line towards Wembley
Park Underground Station

Zone(s): 2, 1

○

Bond Street Underground Station

Take the **Central Line** towards Hainault Underground Station or **Central Line** towards Debden Underground Station or **Central Line** towards Newbury Park Underground Station or **Central Line** towards Epping Underground Station or **Central Line** towards Loughton Underground Station

Arrival

Expected:

Zone(s): 1

17:26

end

Oxford Circus Underground Station

**Average
interchange
time:
4 mins**



Expected journey time: 00:26

Maximum journey time: 00:32

Interchange: 1

Figure 4-18: Proposed modification of the journey planner - Journey with an interchange - Canary Wharf to Oxford Circus

4.6 Conclusion - Assessment of the Proposed Metrics

A set of five journey time reliability metrics are proposed and analyzed in this chapter:

1. The median and the 95th percentile journey time, calculated for each origin-destination pair and timeband, and then aggregated at the line and day levels, give an indication of the central tendency and the spread of the journey time distributions.
2. The difference between the 95th percentile and the median journey times, referred to as the Reliability Buffer Time (RBT) measures the spread of the journey time distribution at the O-D pair and timeband level, and then aggregated at the line and day level.
3. The median and the 95th percentile of the good journey time distribution, defined for each O-D pair and timeband, and aggregated, define standards for comparison of the daily performance.
4. The percentage of passengers traveling in five minutes or more over the good journey time is calculated comparing the actual journey time distribution with the good distribution at the O-D pair and timeband level.
5. The Normalized Reliability Buffer Time (NRBT) normalizes the reliability buffer time by the median journey time to allow the comparison of journey time reliability across lines.

It was proposed to use and publish the five metrics in three different ways. The RBT and the proportion of journey taking 5 minutes (or more) over the good journey time that summarized the five metrics can be published alongside the existing Key Performance Indicators, either every day or at the end of each four-week period. The RBT can also be added to the Journey Planner as proposed in section 4.5.2. The NRBT should be used internally to compare the performance across lines and should not be published with the KPIs. The metrics in their publishable form fulfill the requirements defined in 4.1:

- *Customer-driven.* The journey time reliability metrics defined in this chapter are based on Oyster card data only. Oyster records are produced by the customers and the metrics “give back” to the customers the data they produced in a formatted and aggregated fashion.

- *Simple.* The methodology for the calculation of the Oyster metrics is not as simple as initially expected. However, they are easily computable, and the Oyster records can be processed at the end of every service day to produce the metrics. The RBT is easily understandable for customers if published at the origin-destination pair level on the journey planner. The proportion of journey taking 5 minutes or more over the good journey time, aggregated at the line level, is also easily understandable by customers and the operator.
- *Meaningful for the customers.* The metrics are meaningful for the passengers as they relate to their personal experience. The RBT published on the journey planner also helps the customers plan their trip by acknowledging the inherent variability of journey time.
- *Meaningful for the operator.* The metrics, and especially the proportion of journey taking 5 minutes or more over the good journey time, if published every day alongside the proxy for excess PWT and the LCH can help review each day's performance just after it has happened and take actions to improve future service.
- *Standardizable.* The calculation of the proportion of journey taking 5 minutes or more over the good journey time is based on the calculation of journey time standards. From the standards, targets could be set for every day or every period.

Chapter 5

Measuring Journey Time Reliability for High Frequency Bus Services

As described in chapter 3, London Buses calculates and reports reliability metrics for high frequency bus services that only consider the waiting component of the journey experience: the Excess Wait Time and the percentage chance of waiting more than x minutes. This chapter proposes a framework for measurement of the total journey time reliability for high frequency bus services in London. The goal is to obtain reliability metrics that are similar to the ones previously defined for the London Underground. The metrics must therefore consider the same components of the journey. The bus journey (without a transfer) can be broken into two components: the waiting time at the stop and the in-vehicle travel time¹. These two parts of the the journey experience are affected by the performance of the route used by the customer. The waiting time is taken into account in the calculation of the Oyster reliability metrics defined in chapter 4 as the wait on the platform occurs between the entry and exit validation of the Oyster card. For bus journeys however, the Oyster card validation occurs when the customer boards the bus, and therefore after the wait at the stop. This chapter proposes a method for the estimation of the total journey time that includes both the waiting time at a bus stop, taking into account the destination of the trip, and the in-vehicle travel time.

¹Passenger in-vehicle travel time and bus running time between two stops are used interchangeably in this chapter.

The Bank Station / Princes Street stop for instance is served by four bus routes (routes 21, 43, 76 and 141) as illustrated by Figure 5-1. Depending on their destination, customers waiting at the stop will choose to board different buses. Assuming that they are aware of the possible options and that they can board the first bus that arrives at the stop, customers traveling to Old Street Station can take any of the four routes, while passengers going to Wood Green can only take route 141 for a journey without transfer. The waiting time distributions for these two origin-destination pairs will therefore be different.

The Excess Wait Time (EWT) metric calculated by London Buses at the route level does not consider the situation where passengers may have more than one route to reach their destination. Therefore, the waiting time captured by EWT may overestimate the true waiting time experienced by those customers who have several options available. The methodology developed in this chapter analyzes journey time distributions at the origin-destination pair level in order to deal with this situation.

The proposed methodology is similar to the one developed in chapter 4 for the Underground with the difference that Oyster data allows a direct measurement of the total journey time in the Underground for individual trips, while the journey time on the bus network cannot be directly measured and needs to be inferred from other data sources.

This chapter is organized as follows. Section 5.1 describes the assumptions on which the analysis is based. Section 5.2 proposes a methodology for the journey time analysis at the origin-destination pair and timeband levels. The aggregation at the route level is discussed in section 5.3. Results of the proposed metrics are presented in section 5.4, and section 5.5 discusses the use of the metrics. Section 5.6 gives a brief assessment of the metrics.

5.1 Assumptions

This section lists the assumptions underlying the calculation of the bus journey time metrics. The assumptions listed here are typically valid for high frequency bus routes.

- Each customer is able to board the first bus that arrives that serves his destination. Capacity constraints, that might force customers to wait for the second (or later) vehicle before being able to board a bus, are ignored. This assumption will not always hold. Estimates of load and crowding for each bus can be obtained using the methodology and tools developed in Gordon (2012) (described in chapter 2); however

this would not be sufficient to determine conclusively whether or not a customer is able to board a specific bus. A more detailed analysis would be required, which is beyond the scope of this research.

- Customer arrivals at a stop are random and uniformly distributed during the headway preceding the bus that they boarded. (The headway to consider depends on the origin and the destination of each customer.) This assumption is reasonable for customers traveling on high frequency routes (with a headway of 12 minutes or less), for which they are assumed to show up at the origin stop without checking the schedule beforehand.
- Passengers board the first arriving bus that serves their destination. That assumes that they have perfect knowledge of the bus network. They all know the routes that serve their destination from their origin. In other words, if a bus that is supposed to reach their destination arrives at the stop, the customers would not wait for another bus from another route and would board the arriving vehicle.
- All buses of routes scheduled to serve an origin-destination pair are assumed to serve that origin-destination pair: the effect on customers of buses that are curtailed is therefore ignored for this analysis. The AVL data at the O-D pair level used for the analysis only includes bus trips that reach the destination from the origin. However, the first bus that arrives at the origin stop might be announced as curtailed, and the customers could choose not to board the vehicle if it will not reach their destination. This possibility is not considered here.
- If an origin-destination pair is served by two (or more) bus routes, the in-vehicle travel time is independent of the route that the customer used. This assumption may be violated, for instance if an O-D pair is served by express and local routes, or if two routes can serve the same O-D pair via two different paths. In the first case, customers can, on purpose, decide not to board the first vehicle and wait for the express bus. In the second case, the in-vehicle travel time distributions for the two distinct paths will likely be different and customers could choose one route over the other because of the differences in travel times.

5.2 Methodology at the Origin-Destination Pair Level

The methodology proposed in this section at the origin-destination pair level is similar to the one defined in section 4.2.2 for the London Underground. For each origin-destination pair and timeband, the journey time distribution for each day is compared to that of every other day of the four-week period or quarter. The median journey time and reliability buffer time are calculated as the journey time reliability metrics. However, the analysis of high frequency bus services also includes a measure of waiting time reliability. The journey time reliability calculation for trips on the bus network consists of three steps:

1. A distribution of the waiting time at the stop is estimated, depending on the origin and the destination of the customer. The waiting time reliability metrics are defined from this distribution.
2. A distribution of the in-vehicle travel time is calculated using AVL data, for each origin-destination pair and timeband.
3. The waiting time and the in-vehicle travel time distributions are combined to calculate the total journey time distribution and derive the associated reliability metrics.

Section 5.2.1 describes the data used for the analysis. The waiting time estimation is presented in section 5.2.2; section 5.2.3 mentions the in-vehicle time calculation, and section 5.2.4 describes the total journey time calculation. The definition of journey time reliability standards is discussed in section 5.2.5.

5.2.1 Data

The data used for the bus network differs slightly from the data used for the Underground. Traveling on the bus network does not require an exit validation of the Oyster card, the customers validate their cards only when they board the bus. Therefore, Oyster data does not directly provide information on the customer's destination and in-vehicle travel time. However, it is possible to estimate running times and customer waiting time at the stop using information from the iBus data. The method and tools developed in Gordon (2012) that infer bus journey destination from iBus and Oyster data are used for the aggregation of the metrics. Table 5-1 summarizes the uses of the two data sources.

Data source	Uses
iBus data	<ul style="list-style-type: none"> • Headway calculation at the O-D pair level to estimate the waiting time distribution for a specific time interval • In-vehicle travel time calculation for all the bus trips between the origin and the destination during the time interval of interest • Total journey time distribution for each bus trip calculated as the convolution of the waiting time and the in-vehicle travel time
Oyster data	<ul style="list-style-type: none"> • Bus journeys origin and destination inference to calculate the passenger flow on each O-D pair of the route during the time interval of interest • Aggregation at the route, operator, or network levels, and at the desired temporal level

Table 5-1: Data sources and uses

The waiting time and journey time distributions are estimated using iBus data only and are based only on the assumptions listed in section 5.1. The assumptions made in Gordon (2012) for the bus journey origin and destination inference, and therefore used in this chapter for the calculation of the passenger flow at the O-D level, only influence the aggregation of the metrics.

5.2.2 Waiting Time Estimation

As described previously, iBus data allows the calculation of successive bus arrival times at the stop of interest. The leading headways are calculated as the difference between the arrival time at the stop of bus i and the arrival time of bus $i-1$. The estimation of the waiting time distribution is based on the leading headways. First the waiting time distribution is calculated at the O-D pair and timeband levels; then the waiting time reliability metrics are defined.

Waiting Time Distribution

In this section and the subsequent sections, H_i denotes the leading headways in the timeband and for the O-D pair of interest. The headways are calculated by considering all the routes that serve the O-D pair, and therefore will be different for every pair.

Probability density function. The probability density function (PDF) $f_W(w)$ that a given customer traveling on an origin-destination pair during a specific time period has a waiting time equal to w is given by:

$$f_W(w) = \sum_i [P(\text{arriving in headway } i) \times f_W(w \mid \text{arrived in headway } i)] \quad (5.1)$$

where $P(\text{arriving in headway } i)$ is the probability that the customer arrived during headway H_i and $f(w \mid \text{arrived in headway } i)$ is the probability that the customer has a waiting time equal to w given that he arrived during headway H_i . The sum is calculated over all the headways during the time period of interest.

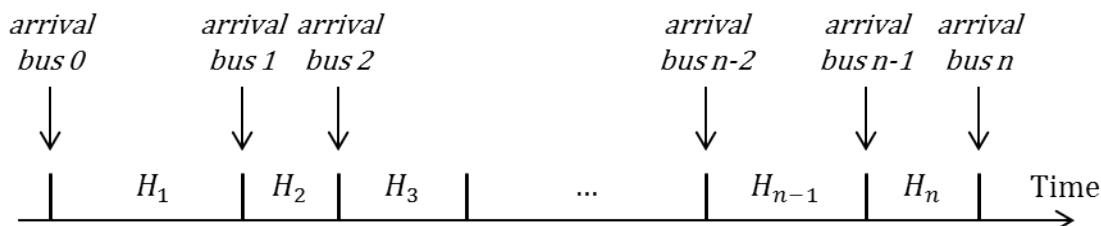


Figure 5-2: Headways and bus arrivals across a time period

Figure 5-2 illustrates the bus arrivals at a given stop across a time period. Passenger arrivals are assumed to be uniformly distributed over this time period. With random passenger arrivals and constant mean arrival rates, the probability of a passenger arriving during headway H_i is proportional to the length of the headway: the longer the headway, the more likely the passenger is to arrive during this interval. Equation 5.2 gives the general form of the probability of the passenger arrival during headway H_i , and therefore the probability of this passenger traveling in bus i .

$$P(\text{arriving in headway } i) = \frac{H_i}{\sum_{i=1}^n H_i} \quad (5.2)$$

Figure 5-3 illustrates the passenger arrival density function over a single headway. Arrivals are random and a passenger traveling on bus i is equally likely to have arrived at any moment during the time interval of length H_i between the arrivals of buses $i - 1$ and i at the stop.

The probability that a passenger waited w , given that he traveled in bus i (or arrived

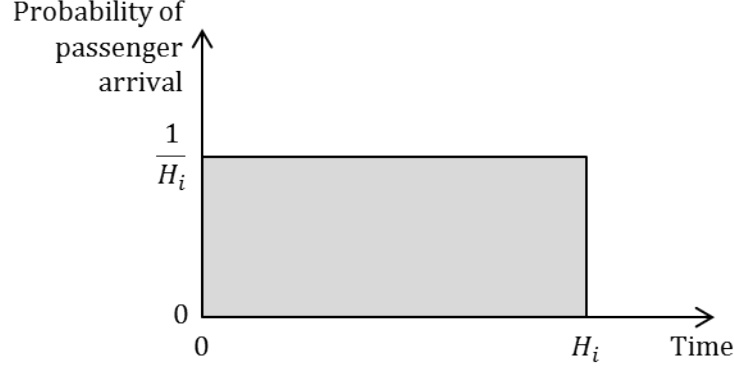


Figure 5-3: Passenger arrival probability density function
(adapted from Ehrlich, 2010)

during headway H_i) is given by Equation 5.3. The probability that a customer waited less than 0 or more than the headway is null.

$$f_W(w \mid \text{arrived in headway } i) = \begin{cases} 0 & \text{if } w < 0 \\ \frac{1}{h_i} & \text{if } 0 \leq w \leq H_i \\ 0 & \text{if } w > H_i \end{cases} \quad (5.3)$$

From Equations 5.2 and 5.3, the probability density function of the waiting time of a passenger is:

$$f_W(w) = \begin{cases} 0 & \text{if } w < 0 \\ \sum_i \left[\frac{H_i}{\sum_{i=1}^n H_i} \times \frac{1}{H_i} \right] & \text{if } 0 \leq w \leq H_i \\ 0 & \text{if } w > H_i \end{cases}$$

$$f_W(w) = \begin{cases} 0 & \text{if } w < 0 \\ \frac{\sum_{i \mid 0 \leq w \leq H_i} 1}{\sum_i H_i} & \text{if } 0 \leq w \leq H_i \\ 0 & \text{if } w > H_i \end{cases} \quad (5.4)$$

Cumulative distribution. The cumulative waiting time distribution $P(W \leq x)$ is calculated by integrating the probability density function from 0 to x . Over the time period of interest, the probability that a customer waits less than x at the bus stop is:

$$P(W \leq x) = \frac{\sum_i (P(W_i \leq x) \times H_i)}{\sum_i H_i} \quad \text{where } P(W_i \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{H_i} & \text{if } 0 \leq x \leq H_i \\ 1 & \text{if } x > H_i \end{cases} \quad (5.5)$$

Waiting Time Reliability

From the waiting time distribution, the median and 95th percentile waiting time are calculated and a *reliability buffer for the waiting time* (noted RBWT) can be estimated for each origin-destination pair and timeband.

The reliability buffer time, or RBT, is defined in section 2.3 as the difference between the 95th percentile and the median of the journey time distribution. Equation 5.6 presents the mathematical formulation of the reliability buffer waiting time for a specific origin-destination pair and timeband.

$$\text{RBWT}_{OD}^t = (WT_{95\%})_{OD}^t - (WT_{50\%}^{\text{ref}})_{OD}^t \quad (5.6)$$

where - t and OD are respectively the timeband and O-D pair of interest

- $(WT_{50\%}^{\text{ref}})_{OD}^t$ is the reference median waiting time for the O-D pair and timeband
- $(WT_{95\%})_{OD}^t$ is the 95th percentile waiting time for the O-D pair and timeband.

The reference median waiting time could be either the median waiting time for the day and timeband of interest or the rolling average median waiting time for the same timeband over a certain number of days. The RBWT is used as an indicator of the waiting time variability, and compared to the reference median waiting time.

5.2.3 In-Vehicle Travel Time

Automatic Vehicle Location (AVL) systems such as iBus allow the calculation of running times for all the bus trips operated during the time period of interest. The in-vehicle travel time for a passenger journey is calculated as the difference between the bus departure time from the origin stop and the arrival time at the destination stop.

Contrary to the waiting time, the in-vehicle travel time is measured directly: for each bus trip, the running time between two stops can be calculated and therefore, if the bus that a passenger boarded is known, his in-vehicle travel time is known as well.

From the list of running times, an in-vehicle travel time distribution can be obtained for each origin-destination pair. Examples of such distributions will be shown in section 5.4. However, bus running times are not the focus of this thesis and will not be analyzed further.

5.2.4 Total Journey Time

The total journey time is calculated as the time from the estimated passenger's arrival at the origin bus stop to the passenger's arrival at the destination bus stop. The journey time considers only the components of the journey over which the bus operator has control: the waiting time at the stop and the in-vehicle travel time. For a given passenger who boarded bus i , the journey time J_i is calculated as the sum of the waiting time W_i and the in-vehicle travel time T_i

$$J_i = W_i + T_i \quad (5.7)$$

The waiting time has to be estimated, given that passengers' arrival times at the bus stop are not known with certainty. The in-vehicle travel time, however, is known, since for every bus trip, AVL data provides travel time from a bus stop to another. As a result, the journey time also needs to be estimated.

Journey Time Distribution

The journey time distribution is calculated for every bus trip between the origin and the destination, and the distribution is then calculated for the entire timeband.

Probability density function. In a similar fashion as for the waiting time distribution, the probability distribution, $f_J(j)$, of the journey time for a given customer traveling on an origin-destination pair during a specific time period is given by:

$$f_J(j) = \sum_i [P(\text{arriving in headway } i) \times f_J(j \mid \text{arrived in headway } i)] \quad (5.8)$$

where $P(\text{arriving in headway } i)$ is the probability that the customer arrived during headway H_i and $f_J(j \mid \text{arrived in headway } i)$ is the probability that the customer has a journey time equal to j given that he arrived during headway H_i . Again, the sum is calculated over all the headways during the time period of interest. The probability that a passenger arrives during headway H_i is given by Equation 5.2.

Figure 5-4 illustrates the journey time distribution for a single bus trip, calculated as the convolution of the distribution of the passenger waiting time and the known in-vehicle travel time.

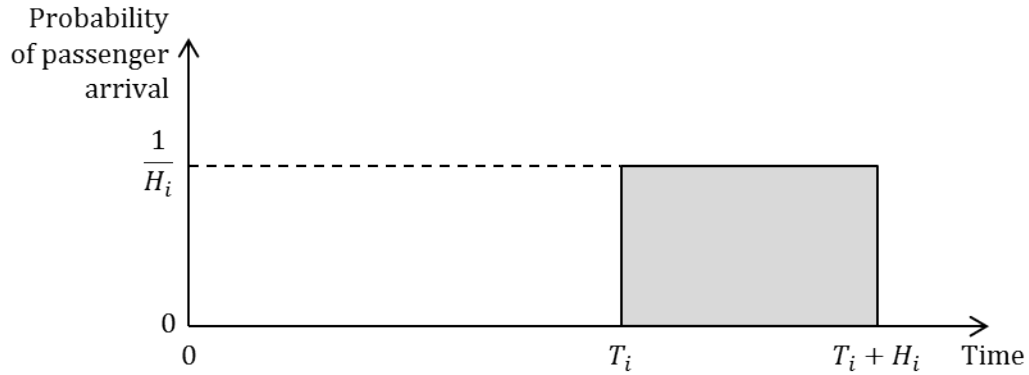


Figure 5-4: Passenger journey time probability density function

For a passenger traveling in vehicle i , his journey time cannot be shorter than the vehicle running time T_i or longer than the sum of the running time T_i and the leading headway H_i . Each passenger traveling in bus i is equally likely to have arrived at any time during headway H_i . The probability that a customer has a journey time equal to j given that he

arrived during headway H_i is calculated as follows:

$$f_J(j \mid \text{arrived in headway } i) = \begin{cases} 0 & \text{if } j < T_i \\ \frac{1}{H_i} & \text{if } T_i \leq j \leq T_i + H_i \\ 0 & \text{if } j > T_i + H_i \end{cases} \quad (5.9)$$

where T_i is the in-vehicle travel time of vehicle i , with H_i the headway preceding vehicle i .

From Equations 5.2 and 5.9, the probability density function of the journey time of a passenger is:

$$f_J(j) = \begin{cases} 0 & \text{if } j < T_i \\ \sum_i \left[\frac{H_i}{\sum_{i=1}^n H_i} \times \frac{1}{H_i} \right] & \text{if } T_i \leq j \leq T_i + H_i \\ 0 & \text{if } j > T_i + H_i \end{cases}$$

$$f_J(j) = \begin{cases} 0 & \text{if } j < T_i \\ \frac{\sum_{i \mid T_i \leq j \leq T_i + H_i} 1}{\sum_i H_i} & \text{if } T_i \leq j \leq T_i + H_i \\ 0 & \text{if } j > T_i + H_i \end{cases} \quad (5.10)$$

Journey Time Reliability

From the journey time distribution, and in a similar fashion as for the waiting time, a *reliability buffer for the total journey time* (referred to as the RBT as for the Underground) is defined for a specific origin-destination pair and timeband as shown by Equation 5.11.

$$\text{RBT}_{OD}^t = (JT_{95\%})_{OD}^t - (JT_{50\%}^{\text{ref}})_{OD}^t \quad (5.11)$$

where - t and OD are respectively the timeband and O-D pair of interest

- $(JT_{50\%}^{\text{ref}})_{OD}^t$ is the reference median journey time for one O-D pair and timeband
- $(JT_{95\%})_{OD}^t$ is the 95th percentile journey time for one O-D pair and timeband.

Similarly to the waiting time, and in a consistent way, the reference median waiting time could be either the median journey time for the day and timeband of interest or the rolling average median journey time for the same timeband over a certain number of days. The RBT, compared the the median journey time, is used as an indicator of journey time reliability at the origin-destination pair and timeband levels.

5.2.5 Waiting Time and Journey Time Reliability Standards

The methodology proposed here is similar to the one described in section 4.2.4 for the Underground. The determination of the good days for each origin-destination pair and each timeband is based on the reliability buffer for the total journey time. A threshold must be defined at the origin-destination pair and timeband levels, between what is considered as a “good” day and what would be called a “bad” day. Once a threshold is defined, an “average” good day can be defined and the journey time distribution for this day is considered as the reference for comparison with every other day. A *percentage of passengers traveling in Δt or more over the good journey time* can be defined as proposed in chapter 4. The threshold Δt can be similar to the one chosen for the Underground to provide a comparable metric.

The time frame over which the reliability standards should be defined depends on the journey time variability. For the Underground, the results showed no evidence of seasonality. However, the bus network is more sensitive to weather than the Underground. Therefore, a correlation between the travel time and the season is likely. The journey time can also vary greatly according to time of day and day of the week, since buses operate in mixed traffic. Consequently, it may be important to define different standards for different days of the week. The determination of journey time reliability standards for the bus network is left for future work.

5.3 Aggregation at the Route Level

The methodology for aggregation at the bus route level is similar to the one developed for the Underground in section 4.2.3. The general formula for aggregation of the journey time reliability metrics at the route level is presented in Equation 5.12. For route R and the time period of interest T (either a timeband or an entire day), a weighted average α^{th} percentile journey time is given by:

$$(JT_{\alpha})_R^T = \frac{\sum_{t \in T} \sum_{OD \in R} (jt_{\alpha})_{OD}^t \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in R} f_{OD}^t} \quad (5.12)$$

- where
- t represents the timebands included in time period T (one timeband if T is a timeband, three is T is an entire day)
 - $OD \in R$ are the origin-destination pairs on line R
 - $(jt_{\alpha})_{OD}^t$ is the α^{th} percentile of the journey time distribution for the origin-destination pair OD and timeband t
 - f_{OD}^t is the passenger flow between origin O and destination D during timeband t .

The aggregated reliability buffer time for route R and time period T is:

$$RBT_R^T = (JT_{95\%})_R^T - (JT_{50\%}^{\text{ref}})_R^T = \frac{\sum_{t \in T} \sum_{OD \in R} \left[(jt_{95\%})_{OD}^t - (jt_{50\%}^{\text{ref}})_{OD}^t \right] \cdot f_{OD}^t}{\sum_{t \in T} \sum_{OD \in R} f_{OD}^t} \quad (5.13)$$

- where
- $(JT_{50\%}^{\text{ref}})_R^T$ is the weighted average median journey time for route R used as a reference
 - $(JT_{95\%})_R^T$ is the weighted average 95th percentile journey time
 - $(jt_{50\%}^{\text{ref}})_{OD}^t$ is the reference median journey time for one O-D pair and timeband
 - $(jt_{95\%})_{OD}^t$ is the 95th percentile journey time for one O-D pair and timeband.

For the Underground, the aggregation at the line level only considered trips that were definitively made on the line of interest. All origin-destination pairs that involve an inter-

change or a path choice were excluded from the data set for aggregation. The aggregation at the route level for the bus network can be performed in a similar fashion. Only trips that are definitively made on the route of interest could be considered. All O-D pairs that are served by more than one route are therefore excluded. Including an O-D pair such as Bank Station to Old Street Station in the aggregation for route 141 could result in attributing to route 141 the poor performance of routes 21, 43 and / or 76, even though these routes are operated by different contractors². This method might exclude from the data set heavy O-D pairs that are served by multiple routes.

Another method would consider all the O-D pairs on the line, including those served by multiple routes. However, the waiting time calculation would be based on route-specific headways, instead of combined headways where all the routes that serve the O-D pair are considered. The method would overestimate the waiting time that passenger experience, but could give a good estimation of the waiting time reliability.

The two methods should be tested and compared.

The weighted average percentiles and buffer time for the waiting time are calculated in a similar fashion. The O-D pair and timeband level median and 95th percentile waiting times are weighted by the passenger flow on the O-D pair of interest. Depending on the methodology chosen for aggregation, the waiting time calculation for an O-D pair served by multiple route would only consider the route-specific headway instead of the combined headway. The passenger flow considered for the calculation of the metrics for a specific route is the volume of passengers using the route of interest only.

²In April 2012, routes 21 and 43 are operated by Go Ahead London and Metroline respectively. Arriva London operates routes 76 and 141.

5.4 Applications

This section illustrates the metrics previously defined through applications at various levels of temporal and spatial aggregation.

Route 141 is used as the application example. Route 141 is a radial route that runs from the London Bridge Bus Station in Central London to the Wood Green Station in the borough of Haringey³ through the City, Hackney, and Islington. The map of the route as of November 2011 is shown in Figure 5-5. The location of the Underground and National Rail stations are also shown along the path.

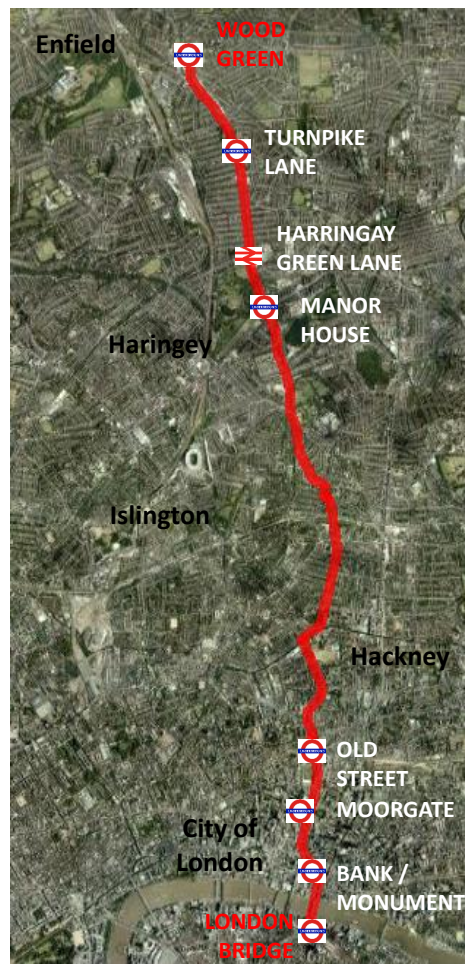


Figure 5-5: Route 141 map as of November 2011 with boroughs and rail stations

The route operates with 5 to 9 minute scheduled headways during the day (AM and PM peaks included) in both directions. Several other bus routes also operate in parallel,

³As of April 2012, the route now runs to Tottenham Road / North Circular Road, north of Wood Green.

reducing the effective headways for passengers whose origin and destination are served by more than one route. Route 141 shares at least two of its stops with 24 other routes⁴. These 24 routes that overlap with 141 will therefore be considered when analyzing the performance of the origin-destination pairs on route 141.

The data available for this application consists of iBus data for all the associated bus routes for the week of 14 - 18 February, 2011, and the week of 17 - 21 October, 2011. The Oyster transaction records for the same two weeks are also available. Only the PM peak in the northbound direction is considered in this application.

5.4.1 Data Processing

Processing the iBus and Oyster data can be tedious and time consuming given the amount available. Oyster data is processed using the methodology defined in Gordon (2012) to calculate the volumes of customers traveling on each origin-destination pair along the route. Sánchez-Martínez (2012) developed and implemented a methodology to calculate headways, running times, and passenger flows at the origin-destination pair level for a specific time interval. The program first reads iBus data and creates three lists of “objects”:

- *A list of “stop visits”*. A “stop visit” includes the route number and direction, the stop code (alphanumeric number), the position of the stop along the route, and the observed arrival and departure times from the stop.
- *A list of “bus trips”*. A “bus trip” includes the route number and direction, and the list of stops visited. From the list of stops, the program determines whether an origin-destination pair is served by the bus trip.
- *A list of “bus stops”*. A “bus stop” includes the stop code, the visiting trips, and the visits at the stop.

From these three lists, for each origin-destination pair on a selected route and in the specified time interval, the program finds the leading headway for each vehicle i , as well as the running time of bus i between the origin and the destination. Figure 5-6 illustrates the methodology.

The leading headway H_i is the time between the arrival of the previous bus $i - 1$ and the arrival of bus i . The running time T_i for trip i is calculated from the departure of bus

⁴The list of stops for route 141 northbound is presented in Appendix D with the list of routes sharing the same physical stops.

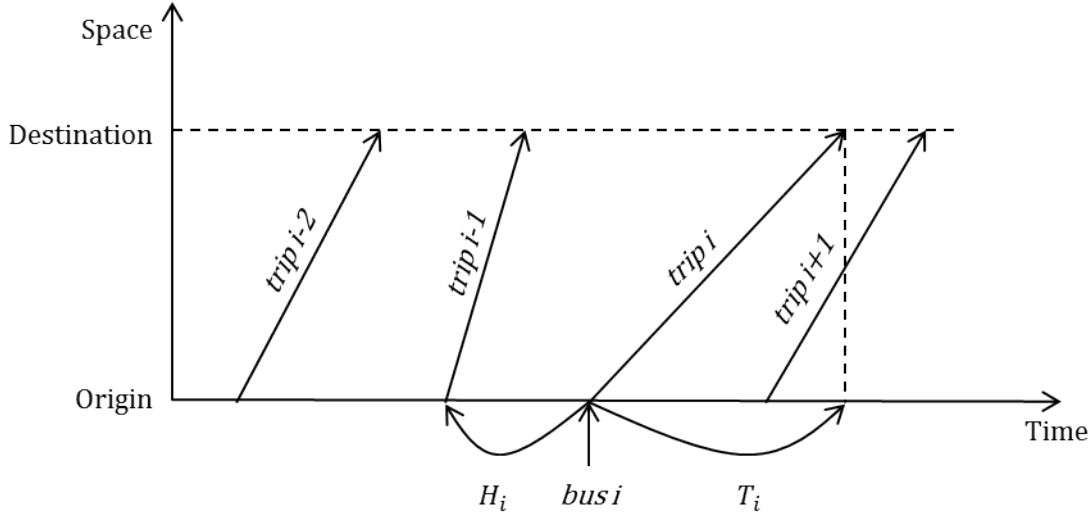


Figure 5-6: Time-space diagram of successive bus trips on a specific origin-destination pair

i from the origin stop to its arrival at the destination stop.

For each O-D pair, the output of the program is a list of leading headways at the origin stop, and the associated running times, as illustrated by Figure 5-7 for four O-D pairs. (The numbers in the O and D columns indicate the stop codes for the origin and destination respectively.)

O	D	H_1	H_2	H_3	...		O	D	T_1	T_2	T_3	...
2468	29071	52	109	1	...		2468	29071	406	306	299	...
2468	862	161	1	62	...	\longleftrightarrow	2468	862	571	573	579	...
2468	2443	1	419	91	...		2468	2443	1181	1145	964	...
2468	30743	420	411	317	...		2468	30743	3072	3002	3140	...

Figure 5-7: Output headways and running times (seconds) for four O-D pairs (adapted from Sánchez-Martínez, 2012)

The tool also reads Oyster data after it is processed using the methodology developed by Gordon (2012) to calculate the passenger flow between origin-destination pairs for the route and the time period of interest.

5.4.2 Results at the Origin-Destination Pair Level

Four origin-destination pairs are used, to illustrate the methodology proposed in this chapter. The O-D pairs share the same origin and are served by one, two, three and four bus routes respectively.

The origin bus stop chosen for the application is the Bank Station / Princes Street stop. From this stop, the four destinations that are considered are the following:

- Old Street Station, served by routes 141, 21, 76 and 43;
- Mintern Street, served by routes 141, 21 and 76;
- Newington Green / Mildmay Road, served by routes 141 and 21;
- Wood Green Station, served only by route 141.

Figure 5-1 shows the four origin-destination pairs on the “spider” bus map for the Bank station area.

These four O-D pairs are chosen because they are served by a different number of routes, and the length of the trip between the origin stop and each destination varies.

Headway Distributions

The study of the headway distribution is not the main objective of this research. However, headway distributions are one of the main building blocks for the calculation of the waiting times. Figure 5-8 illustrates the headway distributions for the four origin-destination pairs of interest over the PM peak of October 17, 2011.

The curve obtained for the Bank Station to Wood Green O-D pair corresponds to the headway distribution for route 141 only. The distribution can be compared to the schedule shown at the bus stop. London Buses advertise a 5 to 8 minute headway at the Bank Station / Princes Street stop for route 141. About 54% of the observed headways are between 5 and 8 minutes, while 27% are longer than 8 minutes and 19% shorter than 5 minutes.

Headways shorter than 60 seconds indicate bunching. For route 141, one route-specific headway was shorter than a minute, as illustrated by the curve for the trip to Wood Green Station. For the other three destinations, combined headways are shorter and bus bunching seems more frequent; this is due to the fact that two (or more) buses serving the destination but operating on different routes might arrive at the origin stop in a short time interval.

Waiting Time Reliability

From the customer’s perspective, the waiting time is more meaningful than the headway distribution, as it relates directly to customers’ experience during their trips. Figure 5-9

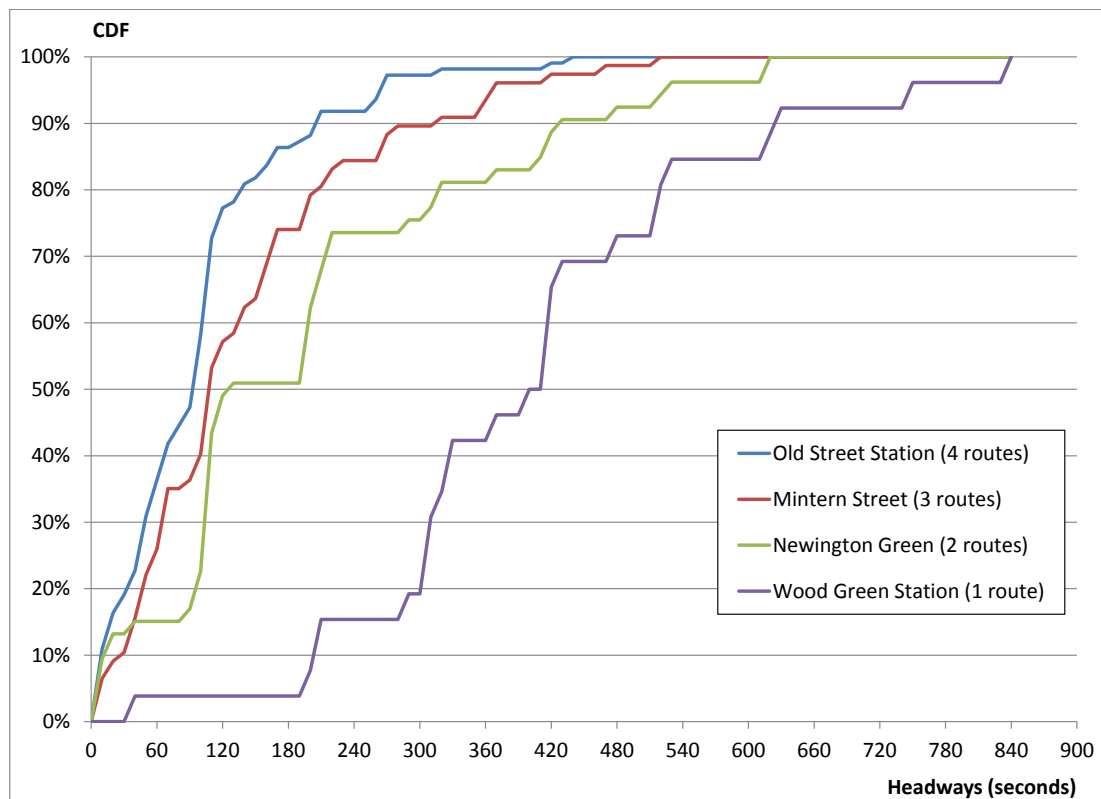


Figure 5-8: Headway cumulative distribution for destinations from the Bank Station / Princes Street stop - October 17, 2011 - PM Peak

illustrates the cumulative distributions of waiting time for the four origin-destination pairs of interest, for the PM peak of October 17, 2011.

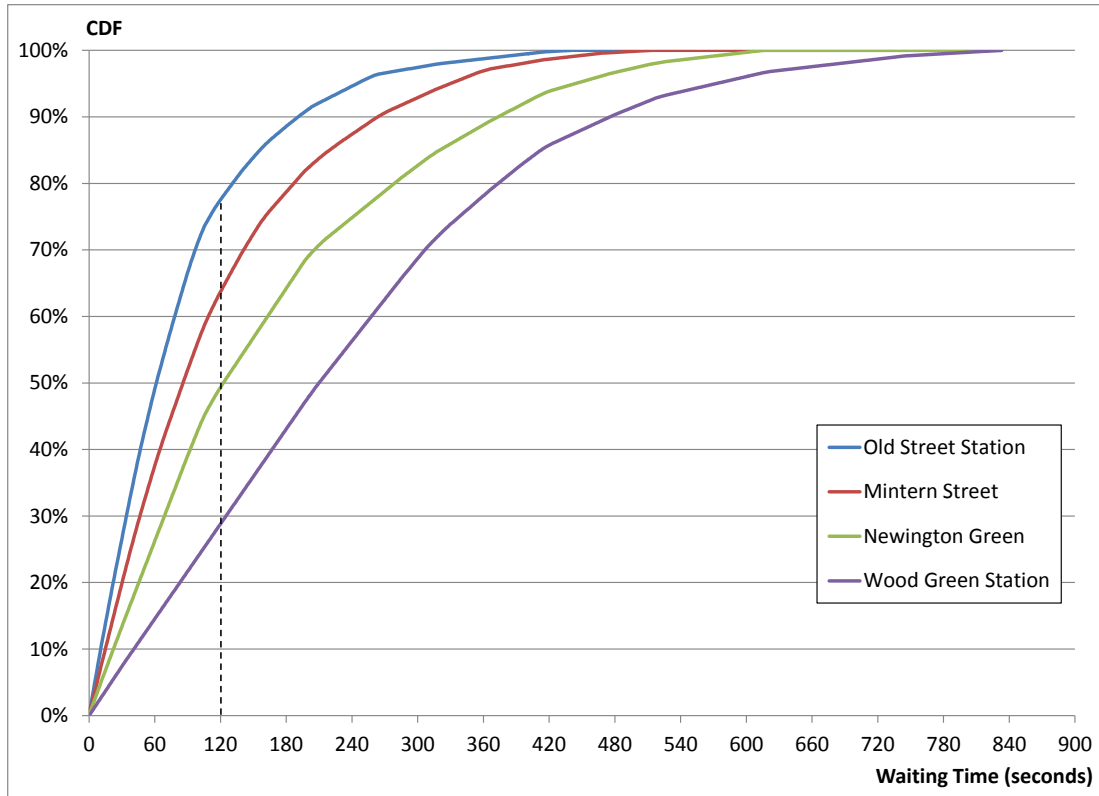


Figure 5-9: Waiting time cumulative distributions for destinations from the Bank Station / Princes Street stop - October 17, 2011 - PM Peak

The waiting time distribution for trips from Bank Station to Wood Green depends only on the headways on route 141. The distributions for the other three O-D pairs take into account the effective headways since passengers may choose any of the route serving their destination. Furthermore, passengers going to Old Street Station will experience a waiting time less than or equal to the waiting time experienced by passengers traveling to Wood Green. These results show that measuring waiting times at QSI points by only considering the headway between consecutive buses on a single route tends to overestimate the actual waiting time that passengers served by more than one route experience.

From the waiting time distributions, the *percentage chance of waiting more than x minutes* (one of the QSI currently used by London Buses) can be calculated for each origin-destination pair and timeband more accurately, and provide a measure of waiting time closer to the customers' experience. For example, only 29% of the customers traveling between

Bank Station / Princes Street and Wood Green Station experienced a waiting time less than or equal to 2 minutes, while more than 77% of the passengers traveling to Old Street Station waited 2 minutes or less.

The waiting time reliability results are shown in Table 5-2 for the weeks of February 14 and October 17, 2011⁵. iBus timestamps are recorded to the second, providing a very high level of precision when compared to the Oyster data. The results presented in Table 5-2 show the relative stability of the median wait times for the four O-D pairs over the two weeks of analysis. The reliability buffer times are twice to three times larger than the median waiting times.

February 14 seems to have been a bad day, with a smaller number of trips, higher median waiting times, and higher reliability buffer times than the other days.

⁵The results for February 16 are not shown in this table as the waiting time distributions calculated for that day were unrealistic, due to errors in the iBus data and / or the data processing.

Day	Old Street Station			Mintern Street			Newington Green			Wood Green Station		
	Bus trips	Median (seconds)	RBWT (seconds)	Bus trips	Median (seconds)	RBWT (seconds)	Bus trips	Median (seconds)	RBWT (seconds)	Bus trips	Median (seconds)	RBWT (seconds)
14-Feb-11	83	80	230	61	108	281	38	153	378	12	540	1078
15-Feb-11	99	62	163	73	88	218	49	119	227	21	245	583
16-Feb-11	—	—	—	—	—	—	—	—	—	—	—	—
17-Feb-11	100	65	175	71	95	217	50	142	324	22	276	547
18-Feb-11	105	60	145	74	89	211	50	135	303	24	247	663
17-Oct-11	110	62	183	77	86	244	53	123	321	26	210	363
18-Oct-11	106	66	147	78	88	192	53	126	265	26	219	495
19-Oct-11	105	64	143	76	87	239	50	125	322	20	271	580
20-Oct-11	104	64	158	72	87	171	50	127	243	24	213	640
21-Oct-11	105	64	136	75	86	217	48	137	358	24	237	469

Table 5-2: Waiting time reliability metrics for destinations from Bank Station / Princes Street - PM peak (16:00 to 19:00)

Journey Time Reliability

Figure 5-10 illustrates the waiting time, in-vehicle travel time, and journey time distributions for the PM peak on October 17, 2011, considering all the bus trips that were made between Bank Station / Princes Street and Old Street Station.

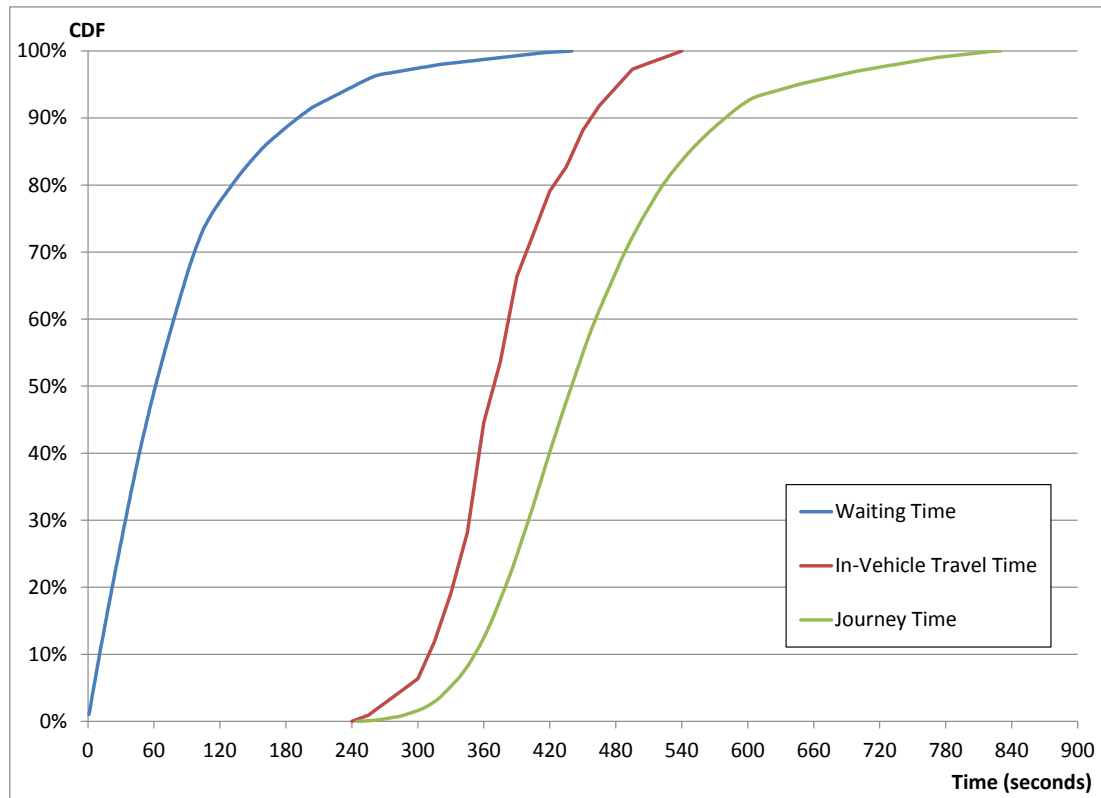
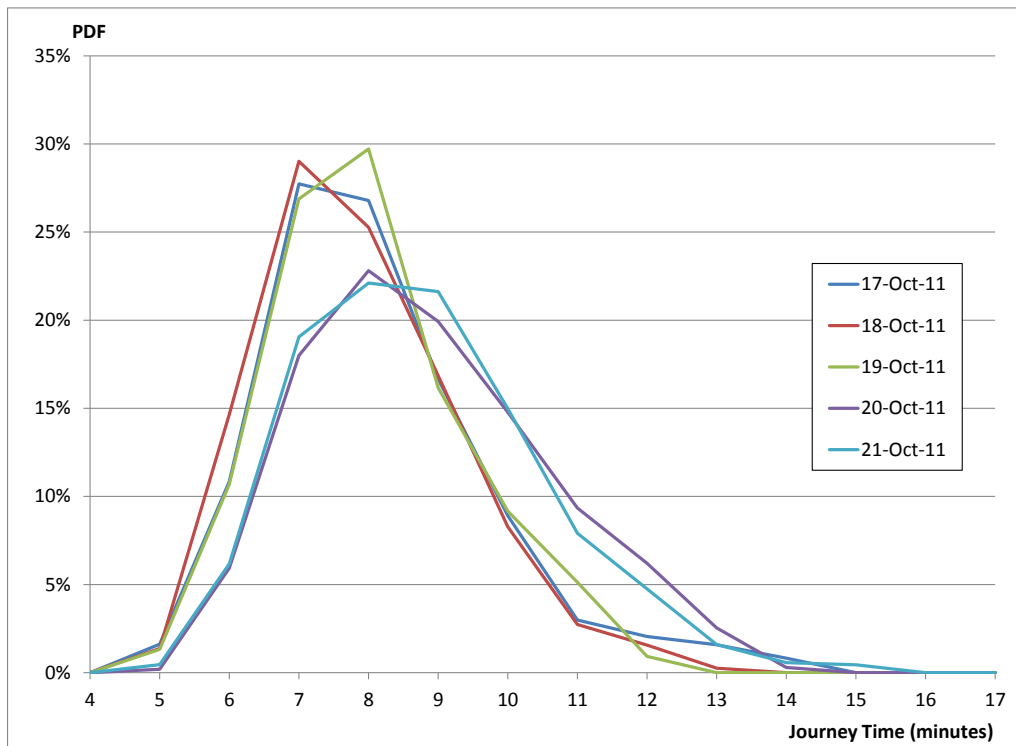


Figure 5-10: Waiting Time, In-vehicle Travel Time, and Journey Time distributions - Bank Station / Princes Street to Old Street Station - October 17, 2011 - PM Peak

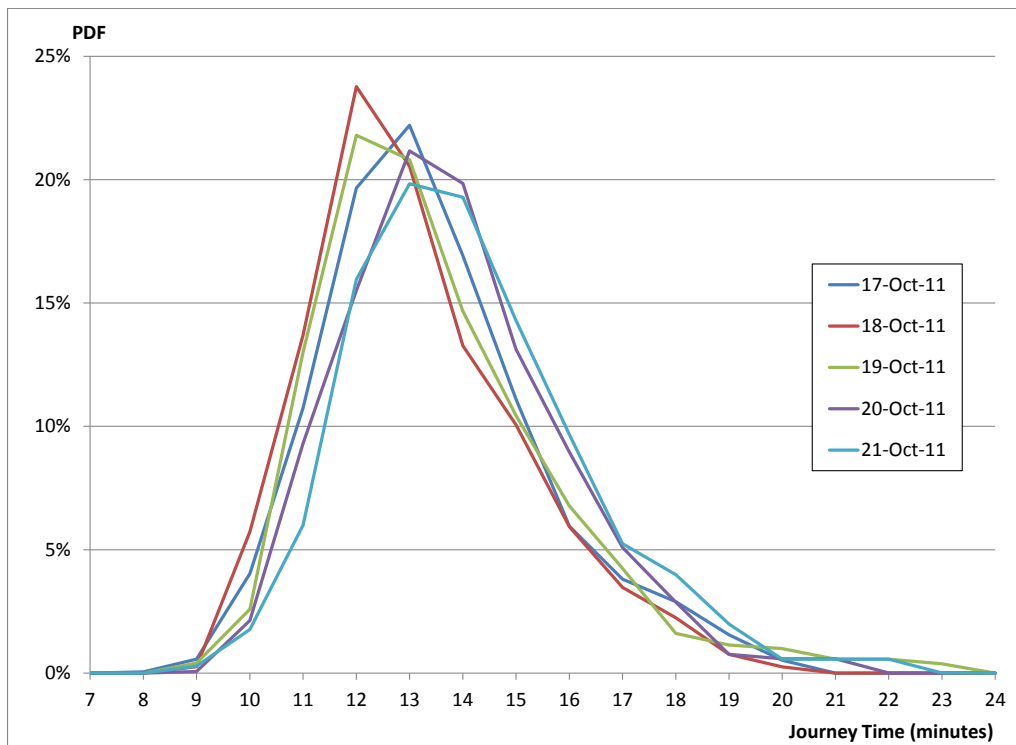
The three distributions in Figure 5-10 show the large variability of the waiting time compared to the in-vehicle travel time, because of the randomness of the passenger arrival process and the headway distributions.

Figure 5-11 compares the journey time distributions, plotted with 1 minute bins, for the four origin-destination pairs of interest, and for the five days in October. For each O-D pair, the distributions appear similar for the five days of the week, except for the trips from Bank Station to Old Street Station on October 20 and 21, and for the trips to Wood Green Station on October 21. The differences could be attributed to incidents on the road network, and a comparison of the bus performance to the traffic incidents could be done.

Table 5-3 shows the journey time reliability metrics for the four O-D pairs and compares

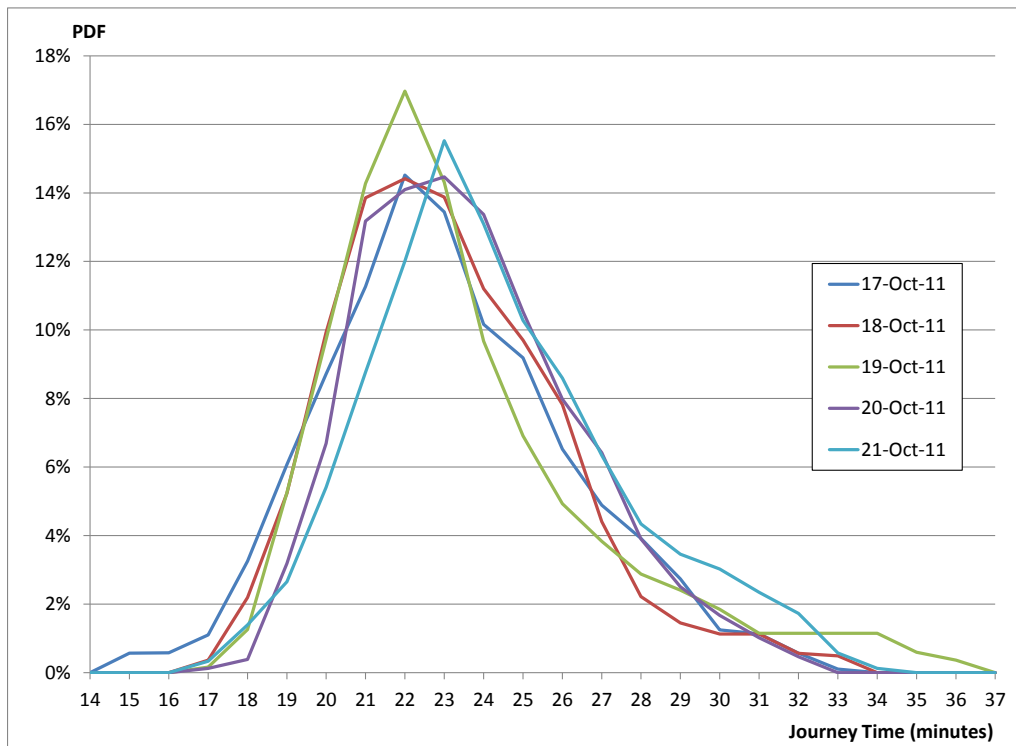


(a) Bank Station / Princes Street to Old Street Station

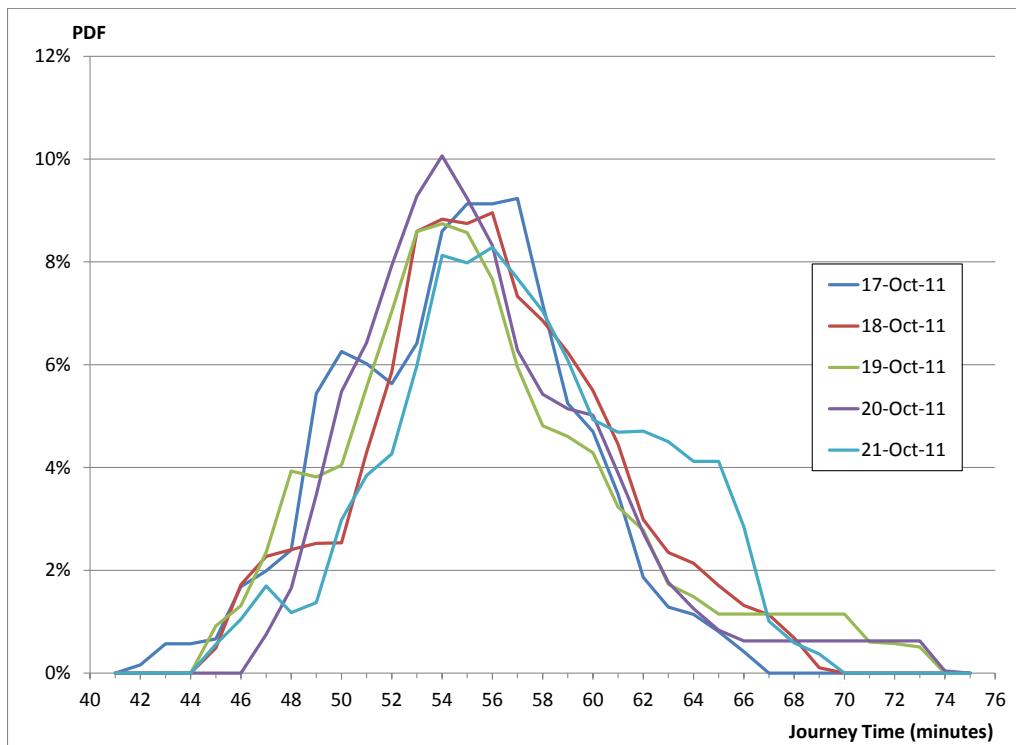


(b) Bank Station / Princes Street to Mintern Street

Figure 5-11: Journey time distributions for 4 O-D Pairs - PM Peak



(c) Bank Station / Princes Street to Newington Green Station



(d) Bank Station / Princes Street to Wood Green Station

Figure 5-11: Journey time distributions for 4 O-D Pairs - PM Peak (Continued)

the performance for the weeks in February and October. The journey time reliability metrics for the week in October are consistent with the distributions presented in Figure 5-11. The performance of the bus routes is consistent and the median journey times and reliability buffer times are similar in October and February. As observed previously with the waiting time metrics, February 14 was a bad day in terms of journey time performance.

Comparison of Waiting Time and Journey Time Reliability

Figure 5-12 compares the waiting time variability to the variability of the journey time for the trips from Bank Station to Wood Green Station. It was previously observed that the waiting time was more variable than the journey time. These two graphs illustrate the difference. Due to the headway variability, the reliability buffer waiting time is proportionally larger than the reliability buffer time. The RBWT is on average twice the median waiting time, while the RBT is only 20% of the median journey time.

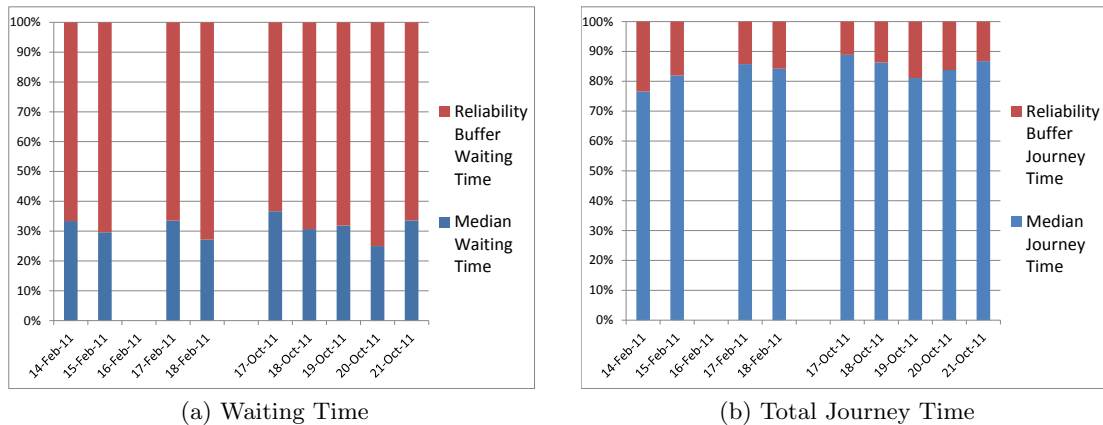


Figure 5-12: Reliability metrics - Bank Station / Princes Street to Wood Green Station - PM peak

Waiting time at bus stops is considered to be one of the most critical components of the total journey time and is heavily weighted in mode choice studies, calculation of level of service, etc. It is also the most variable part of the journey, as illustrated by Figure 5-12. The proposed metrics capture the impact of this variability. For example, particularly bad days, like February 14, impact the reliability metrics.

	Old Street Station			Mintern Street			Newington Green			Wood Green Station		
Day	Bus trips	Median (seconds)	RBT (seconds)	Bus trips	Median (seconds)	RBT (seconds)	Bus trips	Median (seconds)	RBT (seconds)	Bus trips	Median (seconds)	RBT (seconds)
14-Feb-11	83	493	218	61	817	289	38	1432	530	12	4008	1226
15-Feb-11	99	423	163	73	743	248	49	1316	286	21	3458	763
16-Feb-11	—	—	—	—	—	—	—	—	—	—	—	—
17-Feb-11	100	435	185	71	775	254	50	1377	341	22	3379	561
18-Feb-11	105	403	197	74	733	261	50	1326	372	24	3466	650
17-Oct-11	110	440	207	77	761	259	53	1336	358	26	3264	410
18-Oct-11	106	431	165	78	737	250	53	1337	336	26	3312	526
19-Oct-11	105	440	170	76	752	274	50	1329	501	20	3266	762
20-Oct-11	104	489	204	72	785	231	50	1372	320	24	3272	634
21-Oct-11	105	486	205	75	798	262	48	1397	399	24	3381	517

Table 5-3: Journey time reliability metrics for destinations from Bank Station / Princes Street - PM peak (16:00 to 19:00)

5.4.3 Results at the Route Level

This section presents results for the waiting and total travel time distributions at the route⁶ and timeband level. The aggregate results presented here only considered O-D pairs that are served only by route 141. The set of O-D pairs is restricted even more to the origin-destination pairs traveled by a minimum of five persons each day in order to simplify the calculation. 23 O-D pairs are used for the week in February, and 34 pairs for the week in October. These O-D pairs used in the aggregation analysis cover the entire route.

The aggregate metrics are calculated as described in section 5.3. The results for route 141 northbound, for the weeks in February and October 2011, are shown in Figure 5-13.

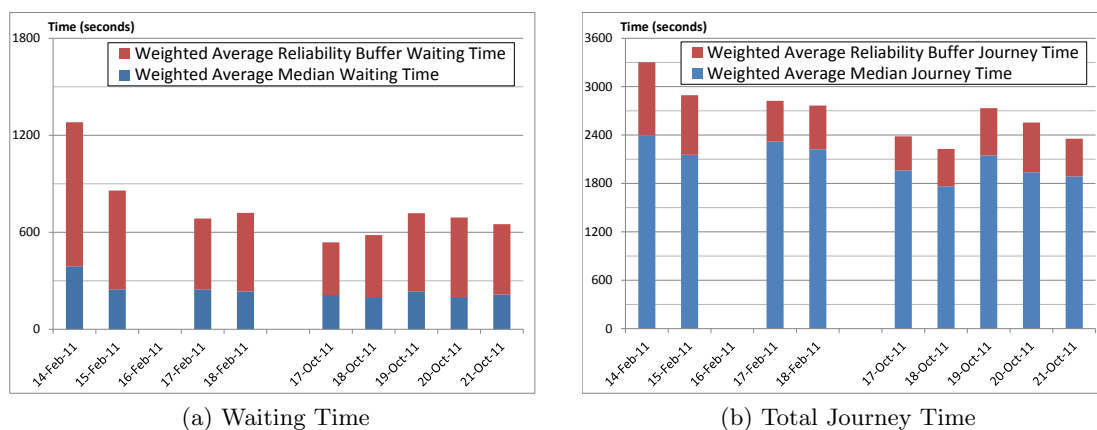


Figure 5-13: Aggregate Reliability Metrics - Route 141 Northbound - PM peak

As observed at the O-D pair level, the performance on February 14 was poor, in terms of both waiting time and total journey time. Considering the other days, the weighted average median waiting time is similar in October and February. The waiting time variability and the weighted average median journey time are higher in February than in October. The journey time variability is similar for the two months.

⁶Results are shown for the northbound direction only.

5.5 Use of the Reliability Metric

The metrics defined in this chapter are intended for both operator and customer uses.

5.5.1 Internal Use at London Buses

The proposed reliability metrics can complement the existing Quality of Service Indicators. The Journey Reliability Buffer Time adds to the existing QSI as the metric takes into account the whole journey experience. The definition of a journey time reliability standard, as mentioned in section 5.2.5, would allow the calculation of the percentage of journeys that take Δt longer than the good journey time, which would add a new customer-focused metric to the set of QSI. This metric would moreover allow comparison with the performance of Underground.

The waiting time reliability metrics might be difficult to publish, as the high reliability buffer waiting time (when compared to the median) can be hard to explain. However, iBus data and the actual headway and waiting time distributions allow the calculation of the Excess Wait Time and the percentage chance of waiting more than x minutes (as defined in chapter 3) using almost the whole population of trips as opposed to the small sample collected through current practice using manual surveys.

The comparison of the proposed reliability metrics with the existing metrics is not possible at this stage, since iBus and Oyster data over an entire quarter are needed to compare the new metrics to the current Quality of Service Indicators that are calculated quarterly.

5.5.2 Journey Planner Modification

In a similar fashion as for the Underground, it is proposed to use the Journey Planner on the TfL website as a medium to report the reliability metrics to the customers, in order to help them plan future journeys.

Figure 5-14 shows the results from the current Journey Planner for trips from Bank Station / Princes Street to Mintern Street, in the PM peak. The Journey Planner takes into account the three possible routes that can be used to reach Mintern Street from Bank Station, and displays information based on the bus timetables. The Journey Planner proposes a first departure at 17:01 (not shown in Figure 5-14), a second at 17:09 and a third at 17:16;



Journey Summary		
Departing: Monday 23 April 2012 at: 17:00 From: Bank Station / Princes Street To: Mintern Street Restrictions:		
Time	Details	Information
17:09	<div>  <div> start </div> </div> Bank Station / Princes Street Take the Route Bus 141 from Stop: A towards Palmers Green or Route Bus 21 from Stop: A towards Newington Green or Route Bus 76 from Stop: A towards Tottenham	Buses every: 2 - 7 mins Max journey time: 17 mins
17:24	<div> <div> end </div> </div> Mintern Street	
Maximum journey time: 00:15 Interchanges: 0		
Time	Details	Information
17:16	<div>  <div> start </div> </div> Bank Station / Princes Street Take the Route Bus 141 from Stop: A towards Palmers Green or Route Bus 76 from Stop: A towards Tottenham or Route Bus 21 from Stop: A towards Newington Green	Buses every: 2 - 6 mins Max journey time: 16 mins
17:31	<div> <div> end </div> </div> Mintern Street	
Maximum journey time: 00:15 Interchanges: 0		

Figure 5-14: Journey Planner - Bus trip from Bank Station / Princes Street to Mintern Street

it also states that the schedule headway is 2 to 6 (or 7) minutes. The 8-minute interval between the first and second proposed departure, followed by a 7 minute interval between the second and third departure are not consistent with the maximum headway reported. The headways reported for the same period are also inconsistent given that the scheduled headways posted at the bus stops do not change within a one hour period. Finally, For the two possible journeys shown in Figure 5-14, three values for the maximum journey time are reported.

Figure 5-15 proposes a possible modification of the Journey Planner for the trip between Bank Station and Mintern Street.

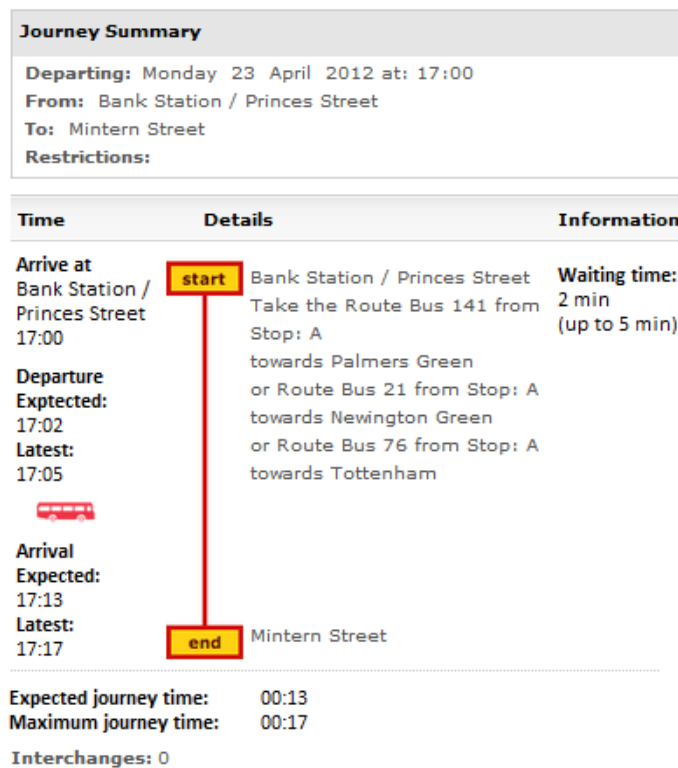


Figure 5-15: Proposed modification of the Journey Planner for bus trips from Bank Station / Princes Street to Mintern Street

For a passenger arrival at the bus stop at 17:00, using results presented in section 5.4, customers should expect a waiting time of 2 minutes, and up to 5 minutes. The expected arrival time would be 17:13 and the maximum arrival time, based on the reliability buffer time, would be 17:17. Displaying a waiting time instead of a range of headways may be more meaningful for customers as it relates to their own experience. An additional modification to the Journey Planner could be considered, in the case when customers select the “Arrive

at” option instead of “Depart at”. The proposed change is similar to the one described in section 4.5.2 for the Underground: the customer arrival time proposed by the Journey Planner is calculated as the requested maximum arrival time at destination minus the 95th percentile of the journey time distribution for the corresponding O-D pair and timeband.

5.6 Conclusion - Assessment of the Proposed Metrics

In order to extend the journey time reliability metrics proposed for the London Underground to high frequency bus services, this chapter analyzed two components of the bus journey experience: the waiting time at the stop and the in-vehicle travel time. From these two components, two sets of metrics were defined at the origin-destination pair and timeband level, then aggregated at the route level:

1. the waiting time reliability metrics, including the median waiting time and the reliability buffer waiting time defined as the difference between the 95th percentile waiting time and the median waiting time;
2. the journey time reliability metrics, including the median journey time and the reliability buffer time defined as the difference between the 95th percentile journey time and the median journey time.

In addition to these two metrics, it was proposed to define reliability standards for the waiting time and journey time in order to compare daily performance with the standards, and calculate a percentage of customers who experience a waiting time or a journey time Δt longer than the standard waiting time or journey time, defined as proposed in section 5.2.5.

The new metrics are valuable for both the operator and the customers. From the operator’s point of view, the proposed waiting time metrics can be used to complement the Excess Wait Time and percentage chance of waiting more than x minutes. The new metrics are based on the iBus data and calculated using almost the entire population, while the current metrics are calculated with manual surveys taken at selected points along each bus route with a limited sample. The journey time metrics take into account the entire journey experience from the passenger’s arrival at his origin stop to his arrival at the destination stop. Therefore, they bring a new perspective to the existing set of Quality of Service

Indicators. From the passenger’s point of view, the methodology proposed at the origin-destination pair level takes into account the entire journey experience (waiting time and in-vehicle travel time), considers all the possible routes that serve the O-D pair, and allows the publication of metrics that relate to individual journeys. The Journey Planner could be modified to show the waiting time and journey time variability in order to help customers plan their trips.

Limitations

The metrics defined in this chapter rely on the assumptions described in section 5.1. The relaxation of these assumptions should be considered for a more accurate representation of the passenger arrival process and calculation of the waiting time.

Additionally, origins and destinations are defined by the code of the physical stop at which passengers wait for their bus. However, two bus routes might not serve the same physical stop but serve the same area. Customers could choose either one of the routes to reach their destination. The situation is best exemplified by the two cases along route 141, illustrated by Figure 5-16. Figure 5-16a shows the intersection of Southgate Road and



Figure 5-16: Stop areas

Englefield Road at the border of the boroughs of Hackney and Islington. Routes 141 and 21 run on Southgate Road and stop at the same location, while route 76 northbound turns on Englefield Road and stops at a distinct point. The two physical stops are around the corner from each other, 65 meters apart. A passenger traveling from Bank Station / Princes

Street to Englefield Road could take any of the three routes and therefore the waiting time calculated for the trips between Bank Station and Englefield Road on route 141 tends to be overestimated, given that route 76 is not considered. Figure 5-16b shows Finsbury Square, near Moorgate. The two stops indicated on the map are 25 meters apart. A customer going to Finsbury Park could take a bus that stops at either of these locations. One could also wait between the two stops and walk to one or the other depending on which bus arrives first.

Chapter 6

Conclusion

This chapter summarizes the research on journey time reliability and proposes ideas for future research. Section 6.1 provides a summary of the work presented in this thesis; section 6.2 highlights the finding from this research and section 6.3 give some suggestions for future work.

6.1 Summary

This thesis proposed an analysis of journey time reliability based on Automated Fare Collection (AFC) and Automatic Vehicle Location (AVL) data, building on previous work by Chan (2007), Uniman (2009) and Ehrlich (2010). The London public transport system provides the background for several applications that demonstrate the proposed methods. The main objective of this research was to define a set of simple, customer-driven reliability metrics, that could be useful and meaningful for both customers and operators. Defining a set of metrics common for the rail and bus networks, despite the differences in automatic fare card validation patterns, was also critical.

The methodology, common to the rail and bus systems, consists first of an analysis of the journey time distribution at the finest spatial and temporal levels available, the origin-destination pair and timeband level. For fully gated systems, such as the London Underground, AFC data provides direct journey time measures for every trip from the fare gate at the entry station to the fare gate at the exit station. For bus networks however, no information is available on passengers' arrival times at the origin bus stop, and thus their waiting times need to be estimated in order to be included in the total journey time

reliability metrics. This additional step of estimating waiting times allows the calculation of waiting time reliability metrics. Furthermore, the methods for estimating waiting times accounts for the multiple overlapping routes that sometimes serve particular O-D pairs.

The proposed metrics are aggregated from the origin-destination pair and timeband level to the line (route) level in order to produce metrics that are more relevant for the operators and can be compared with existing metrics.

Finally, the proposed definition of journey time reliability standards at the origin-destination pair and timeband level evaluates the daily performance of a transit line (route) based on actual observed journey times and not on schedules.

6.2 Findings and Recommendations

The main findings and recommendations from this thesis include:

- **AVL and AFC data allow a customer-focused measurement of journey time variability that is useful and meaningful for both customers and the operator.** Using AVL and AFC data, journey time reliability can be measured at the finest spatial and temporal levels. With such data, journey time distributions can be developed, providing the necessary inputs for metrics that relate to specific journeys. The metrics can then be aggregated at an appropriate level to reflect operator's interests.
- **AVL and AFC based reliability metrics focus on extreme values.** The existing reliability metrics, mainly based on manual surveys and small sample sizes, at best only reflect average performance. As illustrated in this thesis, average performance does not vary much from one day to another, and does not reflect severe disruptions that mainly influence passengers' perception of service quality. AVL and AFC data capture worst performance and allow the calculation of the journey time variability around its average value. The proposed metrics include a measure of extreme delays that the current measures do not consider.
- **Complementarity with existing metrics.** The proposed journey time reliability metrics correlate well with the existing proxy metrics calculated on a daily basis for the Underground. However, the proxy metrics for the Underground, as well as the waiting time metrics for the bus network, do not capture the entire journey experience.

Moreover, the Journey Time Metric and the Excess Wait Time are reported aggregated over a four-week period or a quarter. JTM and EWT can only identify general trends in service performance. The proposed metrics complement the existing metrics: they take into account the entire journey experience and offer a granularity of measurement (by timeband, origin-destination pair) that the existing metrics cannot provide.

- **Calculation of the metrics.** The calculation of the journey time reliability metrics for fully gated systems is straightforward as it is based on AFC records that are readily available. The new metrics are easily computable and can be calculated at the end of every service day, providing an assessment of performance that can be used to improve service. For the bus network, AFC and AVL data need to be processed before calculating the metrics. The algorithms developed by Gordon (2012) and Sánchez-Martínez (2012) provide a good methodology for data processing. Industrialization of processing tools will allow the calculation of the reliability metrics for high frequency bus services in a fast and efficient way.
- **Use the metrics to define journey time reliability standards.** The proposed reliability metrics should be used to define journey time reliability standards by the identification of a representative “good” journey time reliability distribution for each origin-destination pair and timeband. The standards can be used by operators to set service performance target and to give journey time information to the customers at the O-D pair level.
- **Use the metrics to provide information to the customers.** The proposed metrics allow the provision of public information to passenger for two purposes: (1) information about past performance of the system, either at a very disaggregate level (O-D pair and timeband) or at the line and network level; (2) planning of future journeys, based on the journey time reliability standards, and considering the variability of journey time. Journey planners could be used as the medium to publicize journey time variability, measured based on AVL and AFC data, instead of scheduled travel times that do not necessarily reflect what customers experience.

6.3 Future Research

The research presented in this thesis could be extended in several directions. Ideas for future work are summarized in this section, first for fully gated systems, second for the bus network, and third for a possible extension of the reliability metrics to multimodal journeys.

6.3.1 Fully Gated Systems

- **Analysis of travel time reliability at a higher level of resolution.** The analysis conducted in this thesis used the timeband as the minimum level of temporal resolution. An analysis at higher levels of resolution, at the 15-minute interval level for instance, could allow the study of performance within the peak periods, or for special events.
- **Include path choice models in the analysis and aggregation at the line level.** The analysis in this thesis excludes from the line level metrics all origin-destination pairs which involve either a path choice or an interchange between lines. Introduction of path choice models will allow the use of actual levels of crowding and passenger flows for the aggregation at the line level.
- **Measure the number of passengers who choose another path or use another mode in the event of a disruption.** AFC data does not capture journey times if no journeys are made on specific O-D pairs, for instance when a line is closed. When disruptions happen, customers may also adjust their travel plans by changing modes or paths. In the context of this research, attempts were made to measure the volume of passengers that use the line under typical conditions and compare this “typical volume” to the daily passenger flow. However, this analysis did not produce any conclusive results and no metric could be defined that captured well the impact of disruptions on passenger flows.

6.3.2 Bus Networks

Most of the future research recommendations for the bus network are related to the relaxation of the assumptions on passenger arrival and behavior at the bus stops.

- **Relax the assumption of random passenger arrivals at the bus stop.** The assumption that passengers arrive randomly at their origin stop might not hold in every case. In particular, bus stops located at major bus transfer stations (such as Victoria or London Bridge bus stations in London), or in the vicinity of rail stations might see bulk passenger arrivals that coincide with the arrivals of buses (or trains) from other routes. Distinguishing between transferring passengers and passengers starting their journey at the bus stop of interest might also be appropriate. In the first case, an approximation of passenger's arrival at the bus stop can be obtained with the methodology developed by Gordon (2012). In the second case, passenger wait times can only be estimated with the methodology developed in this thesis.
- **Local and express services.** The methodology developed in chapter 5 considers overlapping routes that follow the same path. Origin-destination pairs served by express and local services introduce another complication to the problem as passengers have a choice between waiting for a faster bus or boarding the first vehicle that arrives at the origin stop. AVL and AFC data could facilitate this analysis, as the arrivals of buses at the stop of interest can be known with more certainty, and passenger behavior can also be analyzed. Passenger strategies in choosing between alternative services (e.g. express versus local) are also important to understand better and incorporate in the development of the bus reliability metrics.
- **Impact of real-time information on passengers.** AVL systems allow transit operators to provide real-time information to their customers, either displayed at bus stops or published online and accessible with smart phones. More and more customers take advantage of these new technologies and change their behaviors. Real-time information might affect the assumption of random passenger arrival at bus stops. Further analysis is needed to evaluate the impact of real-time information on passenger arrivals.

- **Compare the distribution of bus arrivals at a stop to the theoretical distribution.** The work described in chapter 5 of this thesis uses the actual distribution of bus arrivals at stops. Developing theoretical models to fit the empirical distributions might help the development of simplified approaches to calculate the waiting time and journey time reliability metrics.
- **Aggregation at the route level.** Two methodologies were proposed in chapter 5 for the aggregation of the reliability metrics at the route level: either by considering the origin-destination pairs served only by the route of interest, or by considering all the O-D pairs on the route and using a route specific headway. The first methodology was illustrated in this thesis, with a dataset limited to the O-D pairs with at least 5 journeys during the PM peak, each day of the week. The impact of excluding O-D pairs that are served by multiple routes should be tested by performing aggregation with the second methodology. Further testing is needed to evaluate the possible approaches.

6.3.3 Extension of the Metrics to Multimodal Journeys

The definition of reliability metrics that could be used for multimodal journeys was the one of initial goals of this work, and one of the motivations for proposing consistent reliability metrics for the bus and the rail networks. The extension of the journey time reliability metrics to multimodal journeys must consider the interchange process between vehicles and modes. Seaborn (2008) and Wang (2010) propose an analysis of transfers between modes from a transit planning perspective. These analyses can be extended to measure reliability of the transfer time.

Four different types of transfers can occur in a transit system. The “arrival stop / station” is the destination of the trip leg before the transfer; the “departure stop / station” is the origin of the trip leg that occurs after the transfer.

- *Rail to rail.* The interchange from rail to rail usually occurs behind the fare gates, and therefore is included in the journey time calculated from the AFC records (for closed systems). In the case of an out-of-station interchange, the customers validate when exiting the arrival station and when entering the departure station. The total interchange time can be measured with Oyster, and so does the total journey time

from origin to destination.

- *Bus to rail.* The transfer time is the difference between the smart card validation time at the rail station (known from the transaction record) and the alighting time at the arrival bus stop (known from AVL data). The interchange time is therefore known, and does not include the platform wait time for the second leg of the journey that occurs inside the rail station.
- *Bus to bus.* The transfer time is the difference between the boarding time in the second bus and the alighting time at the arrival stop, which are known from the AFC and AVL data. The transfer time includes the walking time to the departure stop (null if the first and second bus routes share the same stop) and the waiting time. The transfer time is therefore sensitive to the headways of the second bus boarded (or the combined headways if several routes serve the desired destination).
- *Rail to bus.* The transfer time is calculated from the exit validation time at the arrival rail station and the boarding time in the bus, known from AVL and AFC data. The waiting time at the bus stop is also included in the transfer time.

The four interchange cases are summarized in Table 6-1.

From \ To	Rail	Bus
Rail	<ul style="list-style-type: none"> • Interchange included in the Oyster record, except for an out of station Interchange. • Waiting time included in the AFC record 	<ul style="list-style-type: none"> • Total interchange time measurable from AFC and AVL data. • Waiting time can be estimated but depends on passenger's characteristics and behavior.
Bus	<ul style="list-style-type: none"> • Total interchange time known directly from AFC and AVL records. • Waiting time included in the AFC record for the rail trip. 	<ul style="list-style-type: none"> • Total interchange time known directly from AFC and AVL data. • Waiting time depends on the configuration of the stops, and can be estimated.

Table 6-1: Four interchange cases

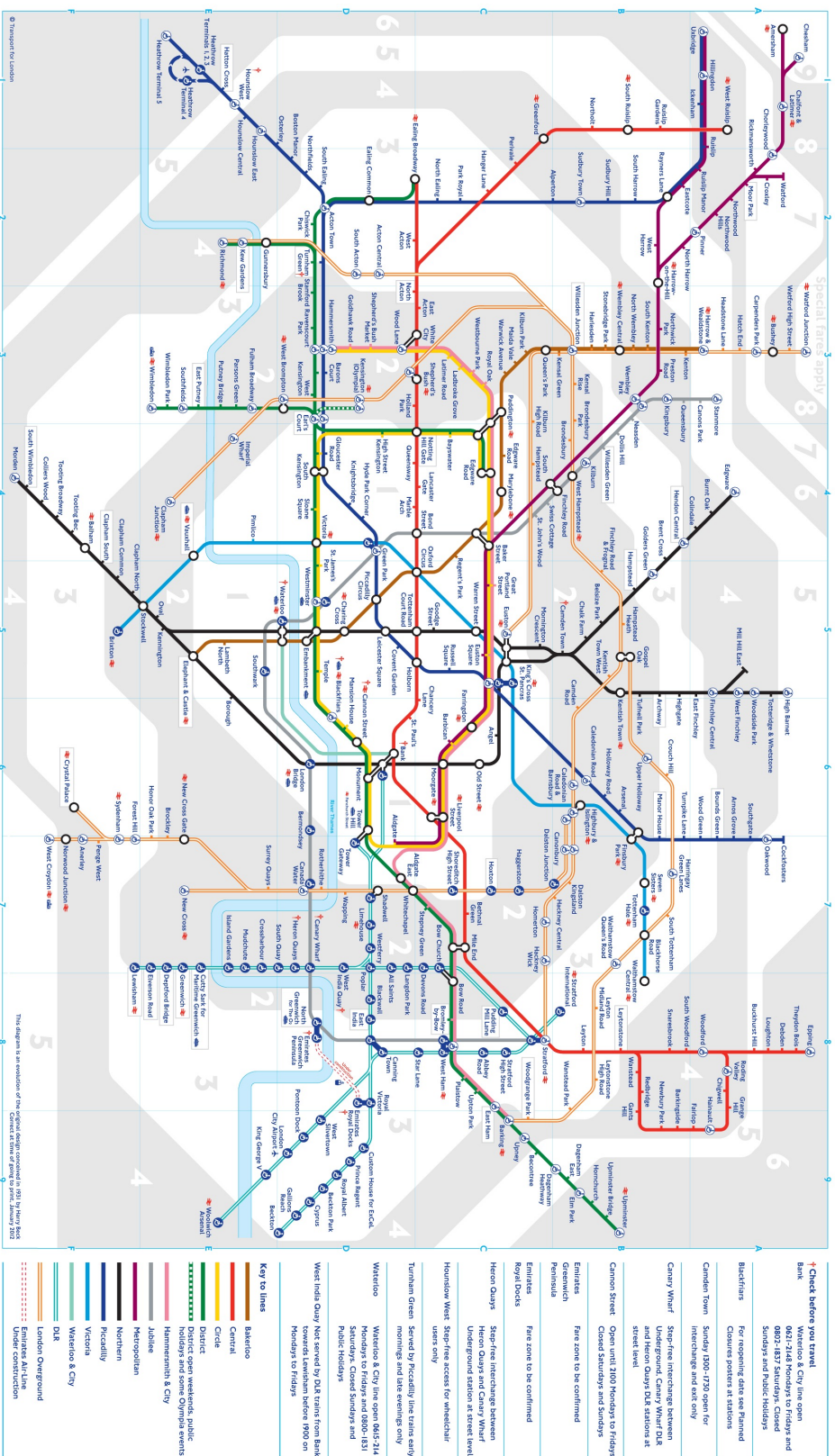
Measuring journey time variability by taking into account all the stages of multimodal journeys would benefit both the customers and the transit agencies. The development of reliability metrics for the total journey time, including transfer time, could help customers plan their future journeys. Considering interchanges and passenger flows on multimodal journeys would also allow transit agencies to improve their operation planning and provide a better quality of service, for instance by designing stations and timing train and bus departures at major transfer points.

Appendix A

Tube Map

Source: <http://www.tfl.gov.uk/assets/downloads/standard-tube-map.gif>.

Tube map



Appendix B

London Underground Reporting Periods

London Underground has thirteen reporting periods per year. Each period is typically 28-day (or four-week) long, but the periods align with the financial year that starts on April 1st and end on March 31st. Therefore the lengths of periods 1 and 13 may vary to align with the Financial Year End of 31 March. The reporting periods for financial year 2011/12 are the following:

Period 1	1 April - 30 April 2011
Period 2	1 May - 28 May 2011
Period 3	29 May - 25 June 2011
Period 4	26 June - 23 July 2011
Period 5	24 July - 20 August 2011
Period 6	21 August - 17 September 2011
Period 7	18 September - 15 October 2011
Period 8	16 October - 12 November 2011
Period 9	13 November - 10 December 2011
Period 10	11 December 2011 - 7 January 2012
Period 11	8 January - 4 February 2012
Period 12	5 February - 3 March 2012
Period 13	4 March - 31 March 2012

Appendix C

London Underground Daily Dashboard

Source: London Underground Intranet (<http://publisher.lul.co.uk/C2/Daily%20Dashboard/default.aspx>). Retrieved on April 30, 2012.

Daily Dashboard...

About Help Reset Print
Next refresh in 11hrs, 22min & 27sec



Day Period to Date Week to Date Last Week

Mon Tue Wed Thu Fri Sat Sun
Selected Date: April 23, 2012

Proxy scores

% of Scheduled Kilometres lost Asset related Customer Hours Excess Platform Wait Time Percentage of Headways Achieved % of Journeys more than 5 min longer than expected

Train cancellations

Morning Evening Station Closures
Cancellations (Scheduled)

Bakerloo	2.0%	2,308	11	98.3%	2.3%	0 (32)	0 (31)	0
Central	6.2%	2,099	23	93.3%	5.0%	0 (75)	2 (76)	0
Circle	See H&C	1,222	See H&C	97.7%	See H&C	1 (31)	0 (32)	0
District	1.1%	4,770	13	97.6%	1.2%	1 (76)	1 (76)	0
Hammersmith & City	4.2%	300	36	94.5%	2.3%	See Circle	See Circle	0
Jubilee	9.1%	6,907	48	90.1%	8.0%	2 (57)	12 (57)	0
Metropolitan	2.6%	5,397	17	95.7%	8.5%	1 (48)	0 (50)	0
Northern	0.6%	5,625	8	99.2%	3.5%	0 (91)	1 (91)	0
Piccadilly	0.6%	106	18	98.3%	1.4%	0 (78)	0 (78)	0
Victoria	21.1%	9,193	0	ND	12.9%	1 (36)	0 (36)	3
Waterloo & City	11.0%	777	28	96.3%		1 (5)	0 (5)	0
All Network	4.8%	38,704				7 (530)	16 (532)	3

Demand and Revenue

Estimated Revenue Actual Comparison to last year (% change)
Estimated Journeys Actual Comparison to last year (% change)

All Network £6,308,029 122.3% 3,605,000 100.0%

Note: ND > No Data captured by core system

Appendix D

Route 141 Northbound

StopCode	StopName	Routes						
35427	LONDON BRIDGE BUS STATION <> #	43						
978	LONDON BRIDGE	21	43	133	149	17	35	
		40	47	48				
2123	MONUMENT STATION <>	21	43	133	149	17	35	
		40	47	48				
27954	BANK STATION <> / KING WILLIAM STREET	21	43	133				
2468	BANK STATION <> / PRINCES STREET	21	43	76				
2467	LONDON WALL	21	43	76				
1160	MOORGATE STATION <> #	21	43	76		205	214	
2463	FINSBURY SQUARE	21		76	271			
29071	OLD STREET STATION <> #	21	43	76	271	205	214	
33436	PROVOST STREET / MOORFIELDS EYE HOSPITAL	21		76	271			
2455	BEVENDEN STREET	21		76	271			
862	MINTERN STREET	21		76	271			
25130	EAGLE WHARF ROAD	21		76	271			
72	BARING STREET	21		76				
2448	DOWNHAM ROAD	21		76				
954	NORTHCHURCH ROAD	21		76				
949	ENGLEFIELD ROAD	21						
13619	BALLS POND ROAD	21						
BP1896	MILDMAY LIBRARY	21						
2443	NEWINGTON GREEN / MILDMAY ROAD	21						
1080	NEWINGTON GREEN		341					
2441	ADEN GROVE		341					
2439	PETHERTON ROAD		341					
2437	GREEN LANES / STOKE NEWINGTON CHURCH ST		341					
BP4326	RIVERSDALE ROAD		341					
2435	KINGS CRESCENT ESTATE		341					
2431	BROWNSWOOD ROAD		341					
BP1202	MYDDLETON AVENUE		341					

StopCode	StopName	Routes					
2429	GLOUCESTER DRIVE		341				
26436	MANOR HOUSE STATION <>		341				
606	MANOR HOUSE STATION <> / WOODBERRY GROVE	29	341				
BP833	ROWLEY GARDENS	29	341				
9993	ENDYMION ROAD	29	341				
4839	HARRINGAY GREEN LANES STATION #	29	341				
9991	MATTISON ROAD	29	341				
1041	ST. ANN'S ROAD	29	341				
9989	BERESFORD ROAD	29					
805	FALKLAND ROAD	29					
29775	TURNPIKE LANE STATION <>	29	230	67			
9984	BRAMPTON PARK ROAD		230	67	144	184	221
9982	WOOD GREEN SHOPPING CITY	29	230	67	144	184	221
		121	123	232	329		
30743	WOOD GREEN STATION <>	29	230	67	144	184	221
		121	123	232	329		

Table D-1: Route 141 Northbound - List of stops and parallel routes

Bibliography

- Abkowitz, M., Slavin, H., Waksman, R., English, L., and Wilson, N. (1978). *Transit Service Reliability*. Report UMTA-MA-06-0049-78-1, USDOT Transportation Systems Center.
- Carrion, C. and Levinson, D. (2012). *Value of travel time reliability: A review of current evidence*. *Transportation Research Part A: Policy and Practice*, volume 46(4): p. 720–741.
- Chan, J. (2007). *Rail Transit O-D Matrix Estimation and Journey Time Reliability Metrics Using Automated Fare Data*. Master’s thesis, Massachusetts Institute of Technology.
- Ehrlich, J.E. (2010). *Applications of Automatic Vehicle Location Systems Towards Improving Service Reliability and Operations Planning in London*. Master’s thesis, Massachusetts Institute of Technology.
- Emmonds, A. and Turner, J. (2010). *Methodology for Generating the Journey Time Reliability Performance Indicator*. London Street Road Network Performance Team.
- Furth, P.G., Hemilly, B., Muller, T.H.J., and Strathman, J.G. (2006). *Using Archived AVL-APC Data to Improve Transit Performance and Management*. TCRP Report 113, Transportation Research Board of the National Academies.
- Furth, P.G. and Muller, T.H. (2006). *Service Reliability and Hidden Waiting Time: Insights from AVL Data*. *Transportation Research Record*, volume 1955: p. 79–87.
- Gordon, J.B. (2012). *Intermodal Passenger Flows on London’s Public Transport Network: Automated Inference of Linked Passenger Journeys Using Fare-Transaction and Vehicle-Location Data*. Master’s thesis, Massachusetts Institute of Technology.
- Lomax, T., Schrank, D., Turner, S., and Margiotta, R. (2003). *Selecting travel reliability measures*. *Texas Transportation Institute monograph*.
- London Buses (2002). *High Frequency Calculations Explained*. Internal London Buses document.
- London Transport (1999). *Journey Time Metric: An Overview*. Market Planning Research Note Series Report M(98)R4.
- MBTA (2012). *ScoreCard January 2012*. http://www.mbta.com/about_the_mbta/scorecard/. Accessed February 22, 2012.
- Muhs, K.J. (2012). *Utilizing Automatically Collected Data to Infer Travel Behavior: A Case Study of the East London Line Extension*. Master’s thesis, Massachusetts Institute of Technology.

- Sánchez-Martínez, G.E. (2012). *Running Time Variability and Resource Allocation: A Data-Driven Analysis of High Frequency Bus Operations*. Master's thesis, Massachusetts Institute of Technology. (forthcoming).
- Seaborn, C.W. (2008). *Application of Smart Card Fare Payment Data to Bus Network Planning in London, UK*. Master's thesis, Massachusetts Institute of Technology.
- STIF (2012). *Les chiffres de la qualité de service - Bulletin d'information trimestriel sur la qualité de service des transports en Île-de-France - Janvier/Septembre 2011*. <http://www.stif.info/IMG/pdf/Les-chiffres-de-la-qualite-2011.pdf>. Accessed February 22, 2012.
- Transport for London (2010). *Transport for London Factsheet*. [http://www.tfl.gov.uk/assets/downloads/corporate/transport-for-london-factsheet\(1\).pdf](http://www.tfl.gov.uk/assets/downloads/corporate/transport-for-london-factsheet(1).pdf). Accessed February 13, 2012.
- Transport for London (2011a). *Developing a reliability metric for LU customers*. Customer research conducted by 2CV for TfL.
- Transport for London (2011b). *London Underground Performance Report, Period 9 2011/12*. <http://www.tfl.gov.uk/assets/downloads/lu-performance-report-period-9-2011-12.pdf>. Accessed February 21, 2012.
- Transport for London (2012a). *London Buses Factsheet*. <http://www.tfl.gov.uk/assets/downloads/corporate/london-buses-factsheet.pdf>. Accessed May 18, 2012.
- Transport for London (2012b). *London Buses Quality of Service Indicators, Network Performance, Third Quarter 2011/12*. <http://www.tfl.gov.uk/assets/downloads/businessandpartners/network-performance-latest-quarter.pdf>. Accessed April 11, 2012.
- Transport for London (2012c). *London Streets Performance Report, Third Quarter 2011/12*. <http://www.tfl.gov.uk/assets/downloads/london-streets-q3-p-report-2011-12.pdf>. Accessed April 14, 2012.
- Transport for London (2012d). *Oyster Factsheet*. <http://www.tfl.gov.uk/assets/downloads/corporate/oyster-factsheet.pdf>. Accessed May 18, 2012.
- Trompet, M., Liu, X., and Graham, D. (2011). *Development of Key Performance Indicator to Compare Regularity of Service Between Urban Bus Operators*. *Transportation Research Record: Journal of the Transportation Research Board*, volume 2216(1): p. 33–41.
- Uniman, D.L. (2009). *Service Reliability Measurement Framework Using Smart Card Data: Application to the London Underground*. Master's thesis, Massachusetts Institute of Technology.
- Uniman, D.L., Attanucci, J., Mishalani, R.G., and Wilson, N.H.M. (2010). *Service Reliability Measurement Using Automated Fare Card Data*. *Transportation Research Record: Journal of the Transportation Research Board*, volume 2143(1): p. 92–99.
- Wang, W. (2010). *Bus Passenger Origin-Destination Estimation and Travel Behavior Using Automated Data Collection Systems in London, UK*. Master's thesis, Massachusetts Institute of Technology.